use ID to match citation to sequence

citations for Swiss Prot & SPTREMBL matches

```
!!AA_SEQUENCE 1.0
ID   O33907      PRELIMINARY;      PRT;   543 AA.
AC   O33907;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Shewanella sp. SCRC-2738.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC   Alteromonadaceae; Shewanella.
OX   NCBI_TaxID=53560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SCRC-2738;
RX   MEDLINE=97419510; PubMed=9274025;
RA   Takeyama H., Takeda D., Yazawa K., Yamada A., Matsunaga T.;
RT   "Expression of the eicosapentaenoic acid synthesis gene cluster from
RT   Shewanella sp. in a transgenic marine cyanobacterium, Synechococcus
RT   sp.";
RL   Microbiology 143:0-0(0).
DR   EMBL; U73935; AAB81126.1; -.
DR   PIR; T30186; T30186.
KW   Hypothetical protein.
SQ   SEQUENCE   543 AA;  59378 MW;  7233F53635B794C7 CRC64;

  O33907  Length: 543  October 5, 2004 14:23  Type: P  Check: 3609  ..

       1  MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRAC YVVNHADHGF

      51  GIAQTADIVT EQAANSTDLP VSAFTPALGT ESLGDNNFRR VHGVKYAYYA

     101  GAMANGISSE ELVIALGQAG ILCGSFGAAG LIPSRVEAAI NRIQAALPNG

     151  PYMFNLIHSP SEPALERGSV ELFLKHKVRT VEASAFLGLT PQIVYYRAAG

     201  LSRDAQGKVV VGNKVIAKVS RTEVAEKFMM PAPAKMLQKL VDDGSITAEQ

     251  MELAQLVPMA DDITAEADSG GHTDNRPLVT LLPTILALKE EIQAKYQYDT

     301  PIRVGCGGGV GTPDAALATF NMGAAYIVTG SINQACVEAG ASDHTRKLLA

     351  TTEMADVTMA PAADMFEMGV KLQVVKRGTL FPMRANKLYE IYTRYDSIEA

     401  IPLDEREKLE KQVFRSSLDE IWAGTVAHFN ERDPKQIERA EGNPKRKMAL

     451  IFRWYLGLSS RWSNSGEVGR EMDYQIWAGP ALGAFNQWAK GSYLDNYQDR

     501  NAVDLAKHLM YGAAYLNRIN SLTAQGVKVP AQLLRWKPNQ RMA

!!AA_SEQUENCE 1.0
ID   Q76636      PRELIMINARY;      PRT;   111 AA.
AC   Q76636;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Vpx protein.
GN   VPX.
OS   Human immunodeficiency virus 2.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11709;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2UC1;
RX   MEDLINE=93124535; PubMed=8419635;
RA   Barnett S.W., Quiroga M., Werner A., Dina D., Levy J.A.;
RT   "Distinguishing features of an infectious molecular clone of the
RT   highly divergent and noncytopathic Human Immunodeficiency Virus type 2
RT   UC1 strain.";
RL   J. Virol. 67:1006-1014(1993).
DR   EMBL; L07625; AAA43944.1; -.
DR   InterPro; IPR000012; RetroV_VpR/X.
DR   Pfam; PF00522; VPR; 1.
KW   AIDS.

SQ   SEQUENCE   111 AA;  12904 MW;  CF152CDB66C0E55C CRC64;

  Q76636  Length: 111  October 5, 2004 14:24  Type: P  Check: 1892  ..

       1  MDPRERVPPG NSDEETIGEA FDWLERTITE LNRVAVNHLP RELIFQVWQR

      51  CWAYWREEQG MSSSYTKYRY LLLMQKAMFV HYTKGCRCLQ EGHGPGGWRS

     101  GPPPPPPPGL A

!!AA_SEQUENCE 1.0
ID   Q7SNM2      PRELIMINARY;      PRT;   111 AA.
AC   Q7SNM2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vpx protein.
GN   VPX.
OS   Human immunodeficiency virus 2.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11709;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=01JP-IMCJ/KR020.1;
RA   Kusagawa S., Imamura Y., Yasuoka A., Oka S., Takebe Y.;
RT   "Identification of HIV-2 subtype B transmission in East Asia.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB100245; BAC79369.1; -.
SQ   SEQUENCE   111 AA;  12786 MW;  06D28ED22BA2B258 CRC64;

  Q7SNM2  Length: 111  October 5, 2004 14:24  Type: P  Check: 973  ..

       1  MDPRERVPPG NSGEETIGEA FEWLERTITE LNRVAVNHLP RELIFQVWQR

      51  SWAYWHEEQG MTISYTKYRY LCLMQKALYM HLARGCACLR EGHGPGGWRP

     101  GPPPPPPPGL A

!!AA_SEQUENCE 1.0
ID   Q7UD35      PRELIMINARY;      PRT;   586 AA.
AC   Q7UD35;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein ycaO.
GN   YCAO OR S0964.
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Shigella.
OX   NCBI_TaxID=623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=22590274; PubMed=12704152;
RA   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA   Schwartz D.C., Blattner F.R.;
RT   "Complete genome sequence and comparative genomics of Shigella
RT   flexneri serotype 2a strain 2457T.";
RL   Infect. Immun. 71:2775-2786(2003).
DR   EMBL; AE016981; AAP16417.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   586 AA;  65667 MW;  1AF2364F0140DBE9 CRC64;

  Q7UD35  Length: 586  October 5, 2004 14:26  Type: P  Check: 8654  ..

       1  MTQTFIPGKD AALEDSIARF QQKLSDLGFQ IEEASWLNPV PNVWSVHIRD

      51  KECALCFTNG KGATKKAALA SALGEYFERL STNYFFADFW LGETIANGPF

     101  VHYPNEKWFP LSENDDVPEG LLDDRLRAFY DPENELTGSM LIDLQSGNED
```

```
     151  RGICGLPFTR QSDNQTVYIP MNIIGNLYVS NGMSAGNTRN EARVQGLSEV

     201  FERYVKNRII AESISLPEIP ADVLARYPAV VEAIETLETE GFPIFAYDGS

     251  LGGQYPVICV VLFNPANGTC FASFGAHPDF GVALERTVTE LLQGRGLKDL

     301  DVFTPPTFDD EEVAEHTNLE THFIDSSGLI SWDLFKQDAD YPFVDWNFSG

     351  TTEEEFATLM AIFNKEDKEV YIADYEHLGV YACRIIVPGM SDIYPAEDLW

     401  LANNSMGSHL RETILSLPGS EWEKEDYLNL IEQLDEEGFD DFTRVRELLG

     451  LATGSDNGWY TLRIGELKAM LALAGGDLEQ ALVWTEWTME FNSSVFSPER

     501  ANYYRCLQTL LLLAQEEDRQ PLQYLNAFVR MYGADAVEAA SAAMSGEAAF

     551  YGLQPVDSDL HAFAAHQSLL KAYEKLQRAK AAFWAK

!!AA_SEQUENCE 1.0
ID   Q7UY73      PRELIMINARY;      PRT;  1080 AA.
AC   Q7UY73;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cation/multidrug efflux pump.
GN   RB832.
OS   Rhodopirellula baltica.
OC   Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC   Planctomycetaceae; Pirellula.
OX   NCBI_TaxID=117;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1;
RX   MEDLINE=22735913; PubMed=12835416;
RA   Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA   Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA   Schlesner H., Amann R., Reinhardt R.;
RT   "Complete genome sequence of the marine planctomycete Pirellula sp.
RT   strain 1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR   EMBL; BX294134; CAD71775.1; -.
KW   Complete proteome.
SQ   SEQUENCE   1080 AA;  118523 MW;  1664093A8ABC8DD8 CRC64;

 Q7UY73  Length: 1080  October 5, 2004 14:26  Type: P  Check: 3909  ..

       1  MKRVLAWAIE NAPGMNVVML ALVIIGAAAF VGMRREVFPE FELEVVMVSV

      51  PYPGATPQDA EEAICQKIEE AIRSIDGIKK VTSIAMEGRG YVLAEMQSDI

     101  QDVQKVMSEI DREVNRIPSF PDLAEDPEIE QITFRDTAIR LGIIGPDDRT

     151  RRGELKLREV AESVRDDLLM LPSVTVAELM GTRNYQIDVE IPEATLRSYG

     201  MTLDQVASEI RQHNVELPGG QLKSSGQEIL LRAKNKGRVG PEIERIPLIT

     251  RQDGVVLTVG DLGNVRDEFE DVTATAEING EPAMVINVQR TKSEDLLNLV

     301  DDVRGYVART EAPPGYRFVL WGDTSVDVRD RMALLLRNGV QGLGLVFLVL

     351  ALFLEVRLAF WVALGIPISI MGAGAALAWG DQTLNMLSLF SFLIALGIVV

     401  DDAIVIGENV YAHRQMGKSL HDAAVDGATE VLPSVAASIT TTVIAFAPMF

     451  FVSGVMGKFM AVIPFAVIAM LLVSLWESTF VLPCHLAHSH SGFFRVATIV

     501  TYPLRPFMLL LFWLNSKASI AMEWFAEKVY VPSLHFCLLN PVLPIAVAIA

     551  LFVGTAGMIR GGVVTTVLFP KSDNNYLQAS VVFPNGTPAS ATEAATKRME

     601  RALQKVSREI ALERAAETRQ PVESLYPPPE ADYLGPVRFA YREVGSITNT

     651  AGPMGGQGSS GSNAGQIFAE LHGTEIREVH SDQLIARWRR EVGEIAGVER

     701  IIYGSIGTGP AGTPIEFKLL ASSEHVDELL AATEVMKQKV GTFAGVFDIS

     751  DDNTPGKWEF QFRVKDKALA TGVTPTDLGQ TVRNTYFGAE VMRLQRGRHE

     801  VKLMVRYPPE ERTSLVNFRE IRVGGADGTQ RPINELAEIN LERGFSEINR

     851  VDQQRSITIS ADLDETTANA DLIIASLQKQ MDGFIAQFPN VSIRWEGQQE

     901  QSRESVGSLM VGFAVAILCM FVLLVLQFRS YMQPLLILAI IPFGMIGAVW

     951  GHAFLDLPLT LFSMFGLVAL AGVVVNDSIV LIDFINSRVR AGDEPIQALL

    1001  ESGRRRFRPI MLTSMTTIAG LIPLLTEKSF QAQLLIPMAS SLAFGLMLAT

    1051  ALVLLLIPVL YMLYLLTLQS LNIPFVEVEE

!!AA_SEQUENCE 1.0
ID   Q7W8I3      PRELIMINARY;      PRT;   267 AA.
AC   Q7W8I3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative amino acid ABC transport system, ATP-binding protein.
GN   BPP2151.
OS   Bordetella parapertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=519;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=12822 / ATCC BAA-587;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640429; CAE37451.1; -.
KW   ATP-binding; Complete proteome.
SQ   SEQUENCE   267 AA;  28615 MW;  46CF6C61E36AE6D0 CRC64;

  Q7W8I3  Length: 267  October 5, 2004 14:25  Type: P  Check: 7664  ..

       1  MKAPTPNTPA LSVRGLAKAF AGVRAVDGLD FDVPAGSVTG LIGPNGCGKS

      51  TSIDCISGFQ PPDAGRVQLA GADIGGLPPH AIARRGLTRT FQNIRLYDEL

     101  TVRENLLIAA QEYDGHGWLD AFLNRRALRD SEAAAGERAR ELIAEVGLSR

     151  YDSLPAGILS YGQKKLVALA ACLMSRPSVA ILDEPLAGVN PTMIANIERI

     201  IDDLNRAGQT FLIVEHNMNF VMRRCHQVVV MESGRKLVEG APSIIRSDER

     251  VVNAYLGSAP QEESHHA

!!AA_SEQUENCE 1.0
ID   Q7WM45      PRELIMINARY;      PRT;   267 AA.
AC   Q7WM45;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative amino acid ABC transport system, ATP-binding protein.
GN   BB1548.
OS   Bordetella bronchiseptica (Alcaligenes bronchisepticus).
```

```
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=518;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB50 / ATCC BAA-588;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640441; CAE32045.1; -.
KW   ATP-binding; Complete proteome.
SQ   SEQUENCE   267 AA;  28615 MW;  46CF6C61E36AE6D0 CRC64;

  Q7WM45  Length: 267  October 5, 2004 14:25  Type: P  Check: 7664  ..

       1  MKAPTPNTPA LSVRGLAKAF AGVRAVDGLD FDVPAGSVTG LIGPNGCGKS

      51  TSIDCISGFQ PPDAGRVQLA GADIGGLPPH AIARRGLTRT FQNIRLYDEL

     101  TVRENLLIAA QEYDGHGWLD AFLNRRALRD SEAAAGERAR ELIAEVGLSR

     151  YDSLPAGILS YGQKKLVALA ACLMSRPSVA ILDEPLAGVN PTMIANIERI

     201  IDDLNRAGQT FLIVEHNMNF VMRRCHQVVV MESGRKLVEG APSIIRSDER

     251  VVNAYLGSAP QEESHHA

!!AA_SEQUENCE 1.0
ID   Q7XXN7      PRELIMINARY;      PRT;   457 AA.
AC   Q7XXN7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome P450.
GN   IHCYP-1.
OS   Iris hollandica (Dutch iris).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Iridaceae;
OC   Iris.
OX   NCBI_TaxID=35876;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Perianth;
RA   Minowa M., Yoshihara N., Imayama T., Tanaka Y., Yabuya T.;
RT   "Molecular cloning of cytochrome P450 from Iris hollandica.";
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB113663; BAC78825.1; -.
SQ   SEQUENCE   457 AA;  51241 MW;  0E58C05666869C3A CRC64;

  Q7XXN7  Length: 457  October 5, 2004 14:23  Type: P  Check: 3447  ..

       1  MVGILVYVAL FILSLLLFLT AAAKLHNSKS KIKNQAPSPP SLPVVGHLHL

      51  LKKPLHRSIS LLSARHGPIL LLRFGSRPAL AVSSLPLAEE CLSGKNDLAF

     101  ANRAHFPHEA APLQLLTLGS ANYGPHWRML RRISAVELLS SHRINSFSQL

     151  RSEEVHSMIS TLFRESSDKE LNRVELKSKL FELAMNNMMR MIFGKDLASS

     201  EGAGRFREMV KESHSLLGAS TRLGDFFPFL GWMDWRARRM VLRLVRRRDE

     251  FLQSLIDAHA RKMEEVEEKT MIRVLVELQK SNRESNNDEG FMLKPLIIGL

     301  LQAGTDTSSD TIEWAMSLLL NNRDKLKKAR DEIDARVGKE RLLRESDLPN

     351  LPYLQCVITE TLRLYPAAPL LVPHESAEEC TVGGYAVPQG TMLLVNAYAI

     401  RVVGIVLGTL IQCFEWERVG EEEVDMTEGS GLTLPRANPL EAICRPRQSM

     451  ISVLAGL

!!AA_SEQUENCE 1.0
ID   Q80Z21      PRELIMINARY;      PRT;  1726 AA.
AC   Q80Z21;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Secreted gel-forming mucin (Fragment).
GN   MUC5AC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RA   Escande F., Buisine M.P.;
RT   "The mouse secreted gel-forming mucin gene cluster.";
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ511870; CAD54412.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR002919; TIL_Cysrich.
DR   InterPro; IPR006552; VC_out.
DR   InterPro; IPR001846; VWF_D.
DR   Pfam; PF01826; TIL; 2.
DR   Pfam; PF00094; vwd; 3.
DR   SMART; SM00215; VWC_out; 2.
DR   SMART; SM00216; VWD; 3.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
FT   NON_TER    1726   1726
SQ   SEQUENCE   1726 AA;  186394 MW;  B90FC8517631057A CRC64;

 Q80Z21  Length: 1726  October 5, 2004 14:24  Type: P  Check: 6085  ..

       1  MGVGRRKLVP FWVLALALAC SQCTGQAQQD SLKSYHEHRS DVPHPQGHVG

      51  TPLNRVTIIP PLKTIPVVRA FNPGHTRRVC STWGNFHYKT FDGQVFYFPG

     101  LCNYVFSAHC GDAYEDFNIQ LRRVQESNTT TLSRVTMKLD GLVVELTKSS

     151  VLVNNHPVQL PFSQSGVLIE LSNGYLKVVA RLGLLFVWNE DDSLLLELDT

     201  KYTNKTCGLC GDFNGSPKSN EFLSNNVRLT PLEFGNLQKM DGPTEQCQDP

     251  LPVPQKNCSA RSGICEMILK GELFSGCAAL VDISSYVEAC RQDVCLCESL

     301  DPSDCICHTL AEYSRQCAHA GGQPQDWRGP NLCSQTCPLN MQHQECGSPC

     351  VDTCSNPQHS QVCEDHCIAG CFCPEGMVLD DINQMGCVPV SQCACLYNGT

     401  LYAPGTNYST DCTNTCSGGQ WSCQDIPCAG TCSVMGGSHM STFDGRQYTV

     451  HGDCTYVLSK PCDSNAFTVL VELRKCGLTE SETCLKTVTL NLGGGQTEIM

     501  VKATGEVFVN QIYTQLPVST ANATFFRPST FFIVGETNLG LQLEIQLSPI

     551  MQTSVRLKPG LRGLTCGLCG NFNSMQADDF QTISGVVEGT AAAFFNTFKT

     601  QAACPNVKNI FQDPCSLSVE NEKYAQHWCS LLTNASGPFS QCHATVNPST
```

```
     651  FFSNCMYDTC NCEKSEDCMC AALSSYVRAC AAKGVLLSDW RDGICTKPTI

     701  TCPKSMTYQY HISTCQPTCR ALNEKDVTCH VSFIPVDGCT CPKGTFLDDL

     751  GKCVQATSCP CYYKGSTVPN GESVQDSGAI CTCTQGALTC IGGPAPTPVC

     801  DAPMIYFDCH NATPGDTGAG CQKSCHTLDM TCYSSECVPG CVCPNGLVAD

     851  GNGGCVVTED CPCVHNEATY RPGETIQVGC NNCTCENRMW QCTDKPCLAT

     901  CAVYGDGHYI TFDGQRYSFN GDCEYTLLQD NCGGNGSSQD AFRVITENIP

     951  CGTTGTTCSK SIKIFLGNYE LKLSDSKMEV VQKDVGQEPP YFVHQMGNYL

    1001  VVETDIGLVL LWDKKTSIFL RLSPEFKGRV CGLCGNFDDN AINDFTTRSQ

    1051  SVVSDMLEFG NSWKLSPSCP DVLVPKDPCT ANPYRKSWAQ KQCSIINSET

    1101  FSACHAHVEP AKYYEACVND ACACDSGGDC ECFCTTVAAY AQACHEVGVC

    1151  VSWRTPDICP LFCDYYNPEG QCEWHYQPCG APCMRTCQNP TGQCLQDLRG

    1201  LEGCYPKCPP TAPIFDEGTM QCVSNCTVTF PCRVNGKLYR PGASVPSDKN

    1251  CDSCICTESG VRCTHNAGAC VCTYNGQQFH PGEIIYHTTD GIGGCISAHC

    1301  RANGTIERSV DTCNSTTPTP PTTFSFSTPP VMTSMQPSST HSSPTPSVGS

    1351  SGASSKAAST TSSILSVKSP VTAPMTMSTS ASAVTTSGCR EECLWSPWMD

    1401  VSRPGRGIDS GDFDTLENLR AHGYPICQVP KAVECRAEAS PGVPLPELQQ

    1451  HLECSTTVGL ICYNSDQLSG LCDNYQIKVQ CCTPVSCPTS QTTHVISSSR

    1501  TTNLDNTTSS VPVTSTEHPY SSTVTSGSST HTPGLSPSSS VPSSPTPASS

    1551  TPAPVSSTTV KTTLPITSPT PEPTPAISSV SISTSGSTMP SSETTHECKQ

    1601  ELCNWTNWLD GSYPGSGRNS GDFDTFVNLR SKGYKFCEKP RNVECRAQFF

    1651  PNTPLEELGQ NVTCSREEGL ICLNKNQLPP MCYNYEIRIE CCTVVNNCST

    1701  ASVTTHPTSH GVSTKTETNW TTHVYS
```

```
!!AA_SEQUENCE 1.0
ID   Q83LP5      PRELIMINARY;      PRT;   597 AA.
AC   Q83LP5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Orf, conserved hypothetical protein.
GN   YCAO OR SF0900.
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Shigella.
OX   NCBI_TaxID=623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=301 / Serotype 2a;
RX   MEDLINE=22272406; PubMed=12384590;
RA   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA   Yu J.;
RT   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT   through comparison with genomes of Escherichia coli K12 and O157.";
RL   Nucleic Acids Res. 30:4432-4441(2002).
DR   EMBL; AE015118; AAN42530.1; -.
DR   InterPro; IPR003776; DUF181.
DR   Pfam; PF02624; DUF181; 1.
DR   TIGRFAMs; TIGR00702; TIGR00702; 1.
KW   Hypothetical protein; Complete proteome.
```

```
SQ   SEQUENCE   597 AA;  67019 MW;  65349DA123B2C6B5 CRC64;

  Q83LP5  Length: 597  October 5, 2004 14:25  Type: P  Check: 8590  ..

       1  MRAFRTTKVI FMTQTFIPGK DAALEDSIAR FQQKLSDLGF QIEEASWLNP

      51  VPNVWSVHIR DKECALCFTN GKGATKKAAL ASALGEYFER LSTNYFFADF

     101  WLGETIANGP FVHYPNEKWF PLSENDDVPE GLLDDRLRAF YDPENELTGS

     151  MLIDLQSGNE DRGICGLPFT RQSDNQTVYI PMNIIGNLYV SNGMSAGNTR

     201  NEARVQGLSE VFERYVKNRI IAESISLPEI PADVLARYPA VVEAIETLET

     251  EGFPIFAYDG SLGGQYPVIC VVLFNPANGT CFASFGAHPD FGVALERTVT

     301  ELLQGRGLKD LDVFTPPTFD DEEVAEHTNL ETHFIDSSGL ISWDLFKQDA

     351  DYPFVDWNFS GTTEEEFATL MAIFNKEDKE VYIADYEHLG VYACRIIVPG

     401  MSDIYPAEDL WLANNSMGSH LRETILSLPG SEWEKEDYLN LIEQLDEEGF

     451  DDFTRVRELL GLATGSDNGW YTLRIGELKA MLALAGGDLE QALVWTEWTM

     501  EFNSSVFSPE RANYYRCLQT LLLLAQEEDR QPLQYLNAFV RMYGADAVEA

     551  ASAAMSGEAA FYGLQPVDSD LHAFAAHQSL LKAYEKLQRA KAAFWAK
```

```
!!AA_SEQUENCE 1.0
ID   Q88HC5      PRELIMINARY;      PRT;   528 AA.
AC   Q88HC5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Conserved hypothetical protein.
GN   PP3435.
OS   Pseudomonas putida (strain KT2440).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=160488;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22423060; PubMed=12534463;
RA   Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA   Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA   Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA   Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA   Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA   Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA   Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA   Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA   Fraser C.M.;
RT   "Complete genome sequence and comparative analysis of the
RT   metabolically versatile Pseudomonas putida KT2440.";
RL   Environ. Microbiol. 4:799-808(2002).
DR   EMBL; AE016786; AAN69037.1; -.
DR   TIGR; PP3435; -.
DR   InterPro; IPR001633; EAL.
DR   Pfam; PF00563; EAL; 1.
DR   PROSITE; PS50883; EAL; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   528 AA;  58370 MW;  3F1689A67A092A25 CRC64;

  Q88HC5  Length: 528  October 5, 2004 14:25  Type: P  Check: 1872  ..

       1  MPLTVKRRAA LSWRTLLPWV VGVLPVMCGL AVMNWQVERE MQASSHATTR

      51  QALEHVEHIL DNLSRAANSL LPLTDSPCEQ AQLMLRAQVT RNAFVRSTNL

     101  FRHNNLYCSS LFGEFEEPVN PGDYVDGKLW LMDGNSVTPG HPLLVYRASD

     151  GDHGAITTVD GDHLLTALRL IGPDEELQIR VGDAWMGKDG VVHKGASPAA
```

```
     201  ASANVLIGST RYPFSVQGGY SADKQGQLLR SHYPGLLSLL LILGALAALA

     251  CRWQIRRATS PRAELDRALQ ADEFLPYFQP VVRKGDYRWA GAEVLMRWHH

     301  PREGLVRPDL FIPYAEHSGQ IVAMTRALLT HTAQSLAPYA GLLEDGFHIG

     351  VNITADHCRD LSLLDDCRTF LQHFPPGRVV LTLELTERKL IEPTPVTLEL

     401  FEKLHAMGVM IALDDFGTGQ SSLNYLRQFK VDYLKIDQSF VAMIGADALS

     451  VHILETIIEL SGKLDLGIVA EGVETDLQRD YLAAHGVDFQ QGYLFARPMP

     501  AAQFLEALAT RPGATQLPQD APPEIMRG

!!AA_SEQUENCE 1.0
ID   Q897L2      PRELIMINARY;      PRT;   375 AA.
AC   Q897L2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Aminotransferase cobD.
GN   COBD OR CTC00722.
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR   EMBL; AE015938; AAO35324.1; -.
DR   GO; GO:0016847; F:1-aminocyclopropane-1-carboxylate synthase . . .; IEA.
DR   GO; GO:0008483; F:transaminase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   InterPro; IPR001176; ACC_synthase.
DR   InterPro; IPR004839; Aminotrans_I/II.
DR   Pfam; PF00155; aminotran_1_2; 1.
DR   PRINTS; PR00753; ACCSYNTHASE.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   375 AA;  43601 MW;  1F72F083CD80183A CRC64;

  Q897L2  Length: 375  October 5, 2004 14:25  Type: P  Check: 3002  ..

       1  MMTLRHGGDI YTDGILKGRK LIDFSSNINP LGIPKSFTEN IQEAIDNLNK

      51  YPDIKYRTLK GNIVEYINKS KVLFQRDEKV IENGYIEIVD ENSIVLGNGA

     101  AEIIDLVIST LKRIMIITPS FGEYEENAEK YGCEILYVNS TKNMDYDYES

     151  ILNNLNRVDG IILGNPNNPD GRILKEEEFK PILDYCEKNN KLVIIDEAFI

     201  EFVGDMSKSY VEKIKNYKCL FIIRALTKFY SMPGIRFGFG ISRDKKVIEE

     251  IKSLQNPWNI NAFAEVAAKY VLKDKEYIKK SLQWITEERE TLLEEIRKLG

     301  FIEKAYNSNG NFILCKLKNI DCEELYDLCL KEDIVIRKCD NYKGLDKSFV

     351  RVAIKDRQTN KILIDKLNKI WRDFK

!!AA_SEQUENCE 1.0
ID   Q8FJB9      PRELIMINARY;      PRT;   597 AA.
AC   Q8FJB9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein ycaO.
GN   YCAO OR C1043.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR   EMBL; AE016758; AAN79513.1; -.
DR   InterPro; IPR003776; DUF181.
DR   Pfam; PF02624; DUF181; 1.
DR   TIGRFAMs; TIGR00702; TIGR00702; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   597 AA;  66951 MW;  D5B9AA2AAC00E5FA CRC64;

  Q8FJB9  Length: 597  October 5, 2004 14:24  Type: P  Check: 8041  ..

       1  MCAFRTTKVI FMTQTFIPGK DAALEDSIAR FQQKLSDLGF QIEEASWLNP

      51  VPNVWSVHIR DKECALCFTN GKGATKKAAL ASALGEYFER LSTNYFFADF

     101  WLGETIANGP FVHYPNEKWF PLTENDDVPE GLLDDRLRAF YDPENELTGS

     151  MLIDLQSGNE DRGICGLPFT RQSDNQTIYI PMNIIGNLYV SNGMSAGNTR

     201  NEARVQGLSE VFERYVKNRI IAESISLPEI PADVLARYPA VVEAIETLEA

     251  EGFPIFAYDG SLGGQYPVIC VVLFNPANGT CFASFGAHPD FGVALERTVT

     301  ELLQGRGLKD LDVFTPPTFD DEEVAEHTNL ETHFIDSSGL ISWDLFKQDA

     351  DYPFVDWSFS GTTEEEPATL MAIFKKEDKE VYIADYEHLG VYACRIIVPG

     401  MSDIYPAEDL WLANNSMGSH LRETILSLPG SEWEKEDYLN LIEQLDEEGF

     451  DDFTRVRELL GLATGSDNGW YTLRIGELKA MLALAGGDLE QALVWTEWTM

     501  EFNSSVFSPE RANYYRCLQT LLLLAQEEDR QPLQYLNAFV RMYGADAVEA

     551  ASAAMSGEAA FYGLQPVDSD LHAFAAHQSL LKAYEKLQRA KAAFWAK

!!AA_SEQUENCE 1.0
ID   Q8JVB6      PRELIMINARY;      PRT;   895 AA.
AC   Q8JVB6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative capsid protein.
OS   Helminthosporium victoriae 145S virus.
OC   Viruses; dsRNA viruses; Partitiviridae; Chrysovirus.
OX   NCBI_TaxID=164750;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Soldevila A.I., Havens W.M., Huang S., Ghabrial S.A.;
RT   "Molecular characterization of Helminthosporium victoriae 145S virus
RT   dsRNA segments 2, 3 and 4.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF297177; AAM68954.1; -.
SQ   SEQUENCE   895 AA;  100352 MW;  C333CD670C5C4485 CRC64;

  Q8JVB6  Length: 895  October 5, 2004 14:24  Type: P  Check: 3063  ..

       1  MADMFNAGKY DRRQAIYMAV RAGAPAFKKA VETVQKLEHW DPTKLPNKMA

      51  DGKPMLFKGL QAVESYMERY SASSLDALSE EKNSFGYGIF GDIRRPAMLG
```

```
     101  ANTVGIDVTV EWGSTEVDAE MHNGENRKMI VSTGTTTVRN EHGKDRAGIS

     151  KATGWDRTEC YSMSPSDVQT LCTLIDTGRA GFNKYTRLVK GMLVYLDLLN

     201  NGKQAIKKRI PNMISYDLKS VLGSYQHRDR SYIYCSNPDS PEYRVVLSLM

     251  SEAYPNDDMT CYGVANIPAD GETNVIVVNG SAGSSVNYHV ELTPQMVMAS

     301  ITQYATESGI ADELESALVC ASSLYHNRYL ARVGLPRVVS SIDLIMPMFR

     351  PANEVRCARP SVAKELAISI GKLHQMCMFL TIKDILVAAR GSTKAGFNYS

     401  SVVESYLTTQ EEVVSVMNAS VTPLRLLEMT PQMKWMYKID AEAMQDLDSL

     451  SIFEVFWLCD GGVASVRNGG VMAFKKGVSD MMTDNPYHDI LRKELAKSNV

     501  VFDFSKLPKG NFTIASRYIR DANEVVLPKL EYVTERVLIA RECDYNPHDR

     551  VEHIINNRTR VTLSARKEAM ESGRTKVAWG DIEVERPTGK GSDESESVLR

     601  FGSSDGRQSP GIEMFGLKSE ARERLEKRRS LVSPPPFRQL SESASTRRAE

     651  RLSVSSHGSR RSISVDLESV RSHSVDGDDD KTPTQSQNLR KRFDFSVLQK

     701  AVDEKKMPGS YESTPEKTEP TVTVEKIPGV KSSMGVSEEV ENDKRAIYKA

     751  DIIGSDRING ISAVNFHKLF RERFDEKQVS PSVMARLMAV LKNVGVRVNI

     801  AAMTADEINA LMKMRDGYDR SYRSSESGHN REFREPDGRV LTMEINLRRM

     851  RCDRDKKPLR EADETGLIPD HMARKLGREF FMAHHERDMI LARVV
```

```
!!AA_SEQUENCE 1.0
ID   Q8XEA9      PRELIMINARY;      PRT;   589 AA.
AC   Q8XEA9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Orf, hypothetical protein.
GN   YCAO OR Z1251 OR ECS0988.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005280; AAG55390.1; -.
DR   EMBL; AP002553; BAB34411.1; -.
DR   PIR; B85616; B85616.
DR   PIR; D90752; D90752.
DR   InterPro; IPR003776; DUF181.
DR   Pfam; PF02624; DUF181; 1.
DR   TIGRFAMs; TIGR00702; TIGR00702; 1.
KW   Complete proteome.
SQ   SEQUENCE   589 AA;  66057 MW;  05925E1EF0A38602 CRC64;

  Q8XEA9  Length: 589  October 5, 2004 14:24  Type: P  Check: 2189  ..

       1  MIFMTQTFIP GKDAALEDSI ARFQQKLSDL GFQIEEASWL NPVPNVWSVH

      51  IRDKECALCF TNGKGATKKA ALASALGEYF ERLSTNYFFA DFWLGETIAN

     101  GPFVHYPNEK WFPLTENDDV PEGLLDERLR AFYDPENELT GSMLIDLQSG

     151  NEDRGICGLP FTRQSDNQTV YIPMNIIGNL YVSNGMSAGN TRNEARVQGL

     201  SEVFERYVKN RIIAESISLP EIPADVLARY PAVVEAIETL EAEGFPIFAY

     251  DGSLGGQYPV ICVVLFNPAN GTCFASFGAH PDFGVALERT VTELLQGRGL

     301  KDLDVFTPPT FDDEEVAEHT NLETHFIDSS GLISWDLFKQ DADYPFVDWN

     351  FSGTTEEEFA TLMAIFNKED KEVYIADYEH LGVYACRIIV PGMSDIYPAE

     401  DLWLANNSMG SHLRETILSL PGSEWEKEDY LNLIEQLDEE GFDDFTRVRE

     451  LLGLATGSDN GWYTLRIGEL KAMLALAGGD LEQALVWTEW TMEFNSSVFS

     501  PERANYYRCL QTLLLLAQEE DRQPLQYLNA FVRMYGADAV EAASAAMSGE

     551  AAFYGLQPVD SDLHAFAAHQ SLLKAYEKLQ RAKAAFWAK
```

```
!!AA_SEQUENCE 1.0
ID   Q8Z806      PRELIMINARY;      PRT;   589 AA.
AC   Q8Z806;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein STY0975.
GN   STY0975 OR T1959.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
DR   EMBL; AL627268; CAD05375.1; -.
DR   EMBL; AE016840; AAO69573.1; -.
DR   InterPro; IPR003776; DUF181.
DR   Pfam; PF02624; DUF181; 1.
DR   TIGRFAMs; TIGR00702; TIGR00702; 1.
KW   Hypothetical protein; Complete proteome.
```

```
SQ   SEQUENCE   589 AA;  65836 MW;  90508293BC06DF6E CRC64;

  Q8Z806  Length: 589  October 5, 2004 14:25  Type: P  Check: 3304  ..

        1  MIFMTQTFIP GKDAALEDSI ARFQQKLLDL GFHIEEASWL NPVPNVWSVH

       51  IRDKECALCF TNGKGATKKA ALASALGEYF ERLSTNYFFA DFWLGETIAN

      101  GPFVHYPNEK WFSLTENDDV PEGLLDARLR AFYDPENELT GSQLIDLQSG

      151  NEERGVCGLP FTRQSDNQTV YIPMNIIGNL YVSNGMSAGN TRNEARVQGL

      201  SEVFERYVKN RIIAESISLP EIPAEVMARY PAVMESIATL EAEGFPIFAY

      251  DGSLGGKYPV ICVVLFNPAN GTCFASFGAH PDFGVALERT VTELLQGRGL

      301  KDLDVFTPPT FDDEEVAEHT NLETHFIDSS GLISWDLFKQ DADYPFTDWS

      351  FSGTTEEEFA TLMAIFAAED KEVYIADYEH LGVYACRIIV PGMSDIYPAE

      401  DLWLANNNMG SHLREILLSL PGSAWNKEDY LNLIEQLDEE GFDDFTRVRE

      451  LLGLATGADN GWYTLRVGEL KAMLALAGGD LEQALIWTEW TMEFNSSVFS

      501  PVRANYYRCL QTLLLLSQED ARQPLQYLNA FIKMYGAEAV EAASTALSGE

      551  AAFYGLPAVD HDLQAFPAHQ SLLKAYDKLQ RAKAAYWSK

!!AA_SEQUENCE 1.0
ID   Q8ZQC7      PRELIMINARY;      PRT;   586 AA.
AC   Q8ZQC7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative cytoplasmic protein.
GN   YCAO OR STM0975.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008741; AAL19909.1; -.
DR   InterPro; IPR003776; DUF181.
DR   Pfam; PF02624; DUF181; 1.
DR   TIGRFAMs; TIGR00702; TIGR00702; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   586 AA;  65342 MW;  E3768BD6D1B944E1 CRC64;

  Q8ZQC7  Length: 586  October 5, 2004 14:24  Type: P  Check: 8250  ..

        1  MTQTFIPGKD AALEDSIARF QQKLLDLGFH IEEASWLNPV PNVWSVHIRD

       51  KECALCFTNG KGATKKAALA SALGEYFERL STNYFFADFW LGETVANGPF

      101  VHYPNEKWFP LTENDDVPEG LLDARLRAFY DPENELTGSQ LIDLQSGNEA

      151  RGVCGLPFTR QSDNQTVYIP MNIIGNLYVS NGMSAGNTRN EARVQGLSEV

      201  FERYVKNRII AESISLPEIP AEVMARYPAV MESIATLEAE GFPIFAYDGS

      251  LGGKYPVICV VLFNPANGTC FASFGAHPDF GVALERTVTE LLQGRGLKDL

      301  DVFTPPTFDD EEVAEHTNLE THFIDSSGLI SWDLFKQDAD YPFTDWSFSG

      351  TTEEEFATLM AIFAAEDKEV YIADYEHLGV YACRIIVPGM SDIYPAEDLW

      401  LANNNMGSHL RETLLSLPGS AWNKEDYLNL IEQLDEEGFD DFTRVRELLG

      451  LATGADNGWY TLRVGELKAM LALAGGDLEQ ALIWTEWTME FNSSVFSPTR

      501  ANYYRCLQTL LLLSQEDARQ PLQYLNAFIK MYGAEAVEAA SAALSGEAAF

      551  YGLPAVDHDL QAFPAHQSLL KAYDKLQRAK AAYWSK

!!AA_SEQUENCE 1.0
ID   Q97WG2      PRELIMINARY;      PRT;   220 AA.
AC   Q97WG2;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein SSO2259.
GN   SSO2259.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR   EMBL; AE006830; AAK42425.1; -.
DR   PIR; B90396; B90396.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   220 AA;  24900 MW;  ECC55875E903DF2B CRC64;

  Q97WG2  Length: 220  October 5, 2004 14:26  Type: P  Check: 6793  ..

        1  MTNGEFYALD NLLKDEKLNS LNKILDVINT TDRLGILDVI KGILEDENTI

       51  GKIIGSLTSD DVLELLVNWD KVIKTTKLFI NEDNIYNIQF LINLIDKVRS

      101  KGILDPIIGL LEDEESLGKI INALINDFTL NLINHWEEII NDLSRIDLTN

      151  FKYYTLLVSA TGEALKTENV KPITSIWEIY KLLKDPDIQR GLGVAASVLK

      201  RIGKLYVPDK GLAFEVEKKL

!!AA_SEQUENCE 1.0
ID   Q9ESP3      PRELIMINARY;      PRT;  1851 AA.
AC   Q9ESP3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Mucin (Fragment).
GN   MUC5AC.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar; TISSUE=Stomach;
RA   Oinuma T., Suganuma T.;
RT   "Rat gastric mucin Muc5AC: Sequence of its 5'-region contains
RT   conserved D-domains and two leucine zipper motifs.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; AB042530; BAB17787.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR002919; TIL_Cysrich.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR001846; VWF_D.
DR   Pfam; PF01826; TIL; 3.
DR   Pfam; PF00094; vwd; 3.
DR   SMART; SM00214; VWC; 3.
DR   SMART; SM00216; VWD; 3.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
FT   NON_TER    1851   1851
SQ   SEQUENCE   1851 AA;  199403 MW;  879DE5B54929C52B CRC64;

 Q9ESP3  Length: 1851  October 5, 2004 14:23  Type: P  Check: 4988  ..

       1  MLHSMGVGRR KLAPFWVLAL ALTFNQHTGQ ALEDTRKSHL EHYSDLSQPQ

      51  GHVGTPLNRV TIIPPLKTIP VVRAFNPAHT RRVCSTWGNF HYKTFDGQVF

     101  YFPGLCNYVF SEHCGAAYED FNIQLRRGLE SNSTTLSRVI MKLDGLVVEL

     151  TKSSVLVNNH PVQLPFSQSG VLIELSNGYL KVVARLGLVF MWNDDDSLLL

     201  ELDTKYANKT CGLCGDFNGS PESSEFLSHN VRLTPLEFGN FQKMDGPTEQ

     251  CQDPLPVPQK NCSIRSSICE EILKGQLFSN CAALVDISSY LEACQQDLCL

     301  CESSDPSNCI CHTLAEYSRQ CAHAGGQPQN WRGPNLCPQT CLLNMEYQEC

     351  GSPCVDTCSN PQHSQVCEDH CVAGCFCPEG MVLDDSNQTG CVPVSQCACL

     401  YNGTLYAPGT SYSTDCTKCT CSGGQWSCQE VPCSGTCSVM GGSHISTFDE

     451  RQYTVHGDCS YVLCKPYDSN AFTVLAELRK CGLTESETCL KTVTLNLGGG

     501  KTVITVKATG EVFVNQIYTQ LPVSTANATF FRPSTFFIIG QTNLGLQLEI

     551  QLHPIMQVSV RIAPEFRGLT SGLCGNFNSM QADDFQTISG VVEGTAAAFF

     601  NTFKTQAACP NVKNIFEDPC SLSVENEKYA QHWCSQLTDA NGPFSQCHAT

     651  VNPSTFFSNC MFDTCNCEKS EDCLCAALSS YVRACAAKGV LLSDWREGIC

     701  AKPTITCPKS MTYQYHISTC QPTCRSLSEE DVTCHVNFIP VDGCTCPKGT

     751  FLDDSGKCVQ ATSCPCYYKG SPVPNGESVH DNGAICTCTQ GALTCIGGPV

     801  LTPVCDAPMI YFDCRNATPG DTGAGCQKSC HTLDMTCYSS ECVPGCVCPN

     851  GLVADGNGSC VVAEDCPCVH NEATYRPGET IQVGCNNCTC ENRMWQCTDK

     901  PCLATCAVYG DGHYITFDGQ RYSFSGDCEY TLLQDNCGGN GSSQDAFRVV

     951  TENIPCGTTG TTCSKGIKIF LGSYELKLSD SKMEVVQKGV GQEPPYFVHQ

    1001  MGNYLVVETD IGLVLLWDKK TSIFLRLSPE FKGKVCGLCG NFDDNAINDF

    1051  TTRSQSVVSD MLEFGNSWKL SPSCPDASVS KDPCTANPYR KSWAQKQCSI

    1101  INSAAFSACH AHVEPAKYYE ACVNDACACD SGGDCECFCT AVAAYAQACH

    1151  EVGVCVSWRT PDICPLFCDY YNPEGQCEWH YQPCGAPCMR TCQNPTGQCL

    1201  QDLRGLEGCY PKCPPTAPIF DEGTMQCVSN CTVSPSPCRV NGKLYRPGTP

    1251  IPSDENCYSC VCTESGVNCT HDAGACVCTY NGQRYHPGDT IYHTTDGMGG

    1301  CISAHCRDNG TIERIVDTCS STSPPPPTTF SFSTTLVMTS MQPSSTHSSP

    1351  TPSVVYPGSP SKAVLTASSV SSVKTPETTS VLTTSTSAST LTMPACQEEC

    1401  LWSPWMDISR PGRGIDSGDF DTLENLHAHG YQICPVPKAV ECRAEDNPGV

    1451  PFHALQQHVE CSTTVGLICY NSDQVSGLCD NYQIKIQCCT PINCPTSTGP

    1501  TQTTHLIVSR TSTMEDTTSS VPVTSTEHTY STVASSPSTH TPGPSPSSSV

    1551  PSSSAPARST PTPVSSTTVK TTLPTTSPMP EPTSATSSVS ISTLGSTLAS

    1601  PEITHGCRKE LCNWTDWIDG SYPEPGRSSG DFDTFVNLRA KGYKFCEKPW

    1651  NVECRAQFFP NTPLQELGQD VTCSREVGLI CLNKNQLPPI CYNYEIRIEC

    1701  CTIVNICSTT SATTQPTSHG VSIKTKTNWI TNTYSFSTEN TSGHSTVINT

    1751  KTWVTGSTHT TPQPGTRPTP SIVSTQDTST SSVQTDSTTS SSHTSSPNTG

    1801  RVSTTHTTHT SSPPTGGTSP TSTTHTSSPN TGGTSPTSTT HTSSPXTGGT

    1851  S

!!AA_SEQUENCE 1.0
ID   Q9N4Q8      PRELIMINARY;      PRT;   307 AA.
AC   Q9N4Q8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   Y5H2B.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Nelson D., Courtney L.;
RT   "The sequence of C. elegans cosmid Y5H2B.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006810; AAF59631.1; -.
DR   WormPep; Y5H2B.4; CE21318.
DR   InterPro; IPR002892; DUF40.
DR   Pfam; PF01838; DUF40; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   307 AA;  34954 MW;  56684308BBE8C9A5 CRC64;

  Q9N4Q8  Length: 307  October 5, 2004 14:23  Type: P  Check: 4608  ..

       1  MDNQPIMNVS ALIVTAIGVF SSLFAFAMNI YILKNFERYL FEQLVNAKKK

      51  NDMALFRFRF FLDVVLGSTT AVYLASMGIS AVYIKESTEF ANFVFHIGFF

     101  TSNIGVSRSL ITLAISIERF VAVYFPILFL QKHPFFPNSV IPIIAIGYGV

     151  FEYPVLYSFC NFELNIPYGC VTLSCSLNLC FYQYWTTYKT IVYILTFLST

     201  IVLTCKLLLK IHESQSNNLT RANRLALIDC IVIFLFDFFP ILFEKLVTQK
```

```
     251  NIFTPRNLGS YSTVTKLSGC AVEAFLVYRT VIRKQDFVNE ATTSKVSSIL

     301  VVKKVTS

!!AA_SEQUENCE 1.0
ID   FA10_HUMAN     STANDARD;      PRT;   488 AA.
AC   P00742; Q14340;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Coagulation factor X precursor (EC 3.4.21.6) (Stuart factor).
GN   F10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91216473; PubMed=1902434;
RA   Messier T.L., Pittman D.D., Long G.L., Kaufman R.J., Church W.R.;
RT   "Cloning and expression in COS-1 cells of a full-length cDNA encoding
RT   human coagulation factor X.";
RL   Gene 99:291-294(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87026600; PubMed=3768336;
RA   Leytus S.P., Foster D.C., Kurachi K., Davie E.W.;
RT   "Gene for human factor X: a blood coagulation factor whose gene
RT   organization is essentially identical with that of factor IX and
RT   protein C.";
RL   Biochemistry 25:5098-5102(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Armel T.Z., Carrington D.P., Chung M.-W., Lee K.L.,
RA   Ozuna M., Poel C.L., Toth E.J., Yi Q., Nickerson D.A.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 13-488 FROM N.A.
RX   MEDLINE=85216545; PubMed=2582420;
RA   Fung M.R., Hay C.W., McGillivray R.T.A.;
RT   "Characterization of an almost full-length cDNA coding for human
RT   blood coagulation factor X.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:3591-3595(1985).
RN   [6]
RP   SEQUENCE OF 19-488 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=86221713; PubMed=3011603;
RA   Kaul R.K., Hildebrand B., Roberts S., Jagadeeswaran P.;
RT   "Isolation and characterization of human blood-coagulation factor X
RT   cDNA.";
RL   Gene 41:311-314(1986).
RN   [7]
RP   SEQUENCE OF 41-179.
RX   MEDLINE=83257207; PubMed=6871167;
RA   McMullen B.A., Fujikawa K., Kisiel W., Sasagawa T., Howald W.N.,
RA   Kwa E.Y., Weinstein B.;
RT   "Complete amino acid sequence of the light chain of human blood
RT   coagulation factor X: evidence for identification of residue 63 as
RT   beta-hydroxyaspartic acid.";
RL   Biochemistry 22:2875-2884(1983).
RN   [8]
RP   SEQUENCE OF 115-488 FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Liver;
RX   MEDLINE=84222026; PubMed=6587384;
RA   Leytus S.P., Chung D.W., Kisiel W., Kurachi K., Davie E.W.;
RT   "Characterization of a cDNA coding for human factor X.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:3699-3702(1984).
RN   [9]
RP   SEQUENCE OF 183-234, AND CARBOHYDRATE-LINKAGE SITES.
RX   MEDLINE=94062825; PubMed=8243461;
RA   Inoue K., Morita T.;
RT   "Identification of O-linked oligosaccharide chains in the activation
RT   peptides of blood coagulation factor X. The role of the carbohydrate
RT   moieties in the activation of factor X.";
RL   Eur. J. Biochem. 218:153-163(1993).
RN   [10]
RP   SEQUENCE OF 1-23 FROM N.A.
RX   MEDLINE=90128299; PubMed=2612918;
RA   Jagadeeswaran P., Reddy S.V., Rao K.J., Hamsabhushanam K., Lyman G.;
RT   "Cloning and characterization of the 5' end (exon 1) of the gene
RT   encoding human factor X.";
RL   Gene 84:517-519(1989).
RN   [11]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 86-179 AND 235-278.
RX   MEDLINE=93360277; PubMed=8355279;
RA   Padmanabhan K., Padmanabhan K.P., Tulinsky A., Park C.H., Bode W.,
RA   Huber R., Blankenship D.T., Cardin A.D., Kisiel W.;
RT   "Structure of human des(1-45) factor Xa at 2.2-A resolution.";
RL   J. Mol. Biol. 232:947-966(1993).
RN   [12]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 86-179 AND 235-278.
RX   MEDLINE=98283982; PubMed=9618463;
RA   Kamata K., Kawamoto H., Honma T., Iwama T., Kim S.H.;
RT   "Structural basis for chemical inhibition of human blood coagulation
RT   factor Xa.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:6630-6635(1998).
RN   [13]
RP   VARIANTS ILE-7 AND HIS-30.
RX   MEDLINE=99318093; PubMed=10391209;
RA   Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA   Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RT   "Characterization of single-nucleotide polymorphisms in coding regions
RT   of human genes.";
RL   Nat. Genet. 22:231-238(1999).
RN   [14]
RP   ERRATUM.
RA   Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA   Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RL   Nat. Genet. 23:373-373(1999).
CC   -!- FUNCTION: Factor Xa is a vitamin K-dependent glycoprotein that
CC       converts prothrombin to thrombin in the presence of factor Va,
CC       calcium and phospholipid during blood clotting.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Thr and then
CC       Arg-|-Ile bonds in prothrombin to form thrombin.
CC   -!- SUBUNIT: The two chains are formed from a single-chain precursor
CC       by the excision of two Arg residues and are held together by 1 or
CC       more disulfide bonds.
```

```
CC   -!- TISSUE SPECIFICITY: Plasma; synthesized in the liver.
CC   -!- PTM: The vitamin K-dependent, enzymatic carboxylation of some
CC       glutamate residues allows the modified protein to bind calcium.
CC   -!- PTM: N- and O-glycosylated.
CC   -!- PTM: THE ACTIVATION PEPTIDE IS CLEAVED BY FACTOR IXA (IN THE
CC       INTRINSIC PATHWAY), OR BY FACTOR VIIA (IN THE EXTRINSIC PATHWAY).
CC   -!- SIMILARITY: Belongs to peptidase family S1.
CC   -!- SIMILARITY: Contains 2 EGF-like domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K03194; AAA52490.1; -.
DR   EMBL; M57285; AAA52421.1; -.
DR   EMBL; AF503510; AAM19347.1; -.
DR   EMBL; BC046125; AAH46125.1; -.
DR   EMBL; L29433; AAA52764.1; -.
DR   EMBL; L00390; AAA52764.1; JOINED.
DR   EMBL; L00391; AAA52764.1; JOINED.
DR   EMBL; L00392; AAA52764.1; JOINED.
DR   EMBL; L00393; AAA52764.1; JOINED.
DR   EMBL; L00394; AAA52764.1; JOINED.
DR   EMBL; L00395; AAA52764.1; JOINED.
DR   EMBL; L00396; AAA52764.1; JOINED.
DR   EMBL; M22613; AAA51984.1; -.
DR   EMBL; K01886; AAA52486.1; -.
DR   EMBL; M33297; AAA52636.1; -.
DR   PIR; A24478; EXHU.
DR   PDB; 1HCG; 08-MAY-95.
DR   PDB; 1FAX; 29-OCT-97.
DR   PDB; 1FXY; 17-JUN-98.
DR   PDB; 1XKA; 23-MAR-99.
DR   PDB; 1XKB; 23-MAR-99.
DR   PDB; 1EZQ; 20-SEP-00.
DR   PDB; 1F0R; 20-SEP-00.
DR   PDB; 1F0S; 20-SEP-00.
DR   PDB; 1FJS; 17-NOV-00.
DR   PDB; 1G2L; 20-OCT-01.
DR   PDB; 1G2M; 20-OCT-01.
DR   PDB; 1KSN; 19-JUN-02.
DR   PDB; 1KYE; 11-FEB-03.
DR   PDB; 1MQ5; 28-JAN-03.
DR   PDB; 1MQ6; 28-JAN-03.
DR   PDB; 1NFU; 25-FEB-03.
DR   PDB; 1NFW; 25-FEB-03.
DR   PDB; 1NFX; 25-FEB-03.
DR   PDB; 1NFY; 25-FEB-03.
DR   MEROPS; S01.216; -.
DR   GlycoSuiteDB; P00742; -.
DR   Genew; HGNC:3528; F10.
DR   MIM; 134530; -.
DR   MIM; 227600; -.
DR   GO; GO:0005576; C:extracellular; TAS.
DR   GO; GO:0003804; F:blood coagulation factor X activity; TAS.
DR   GO; GO:0007596; P:blood coagulation; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR001254; Peptidase_S1.
DR   InterPro; IPR001314; Peptidase_S1A.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00001; GLABLOOD.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50026; EGF_3; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Glycoprotein; Hydrolase; Serine protease; Plasma; Blood coagulation;
KW   Gamma-carboxyglutamic acid; Hydroxylation; Calcium-binding; Vitamin K;
KW   Signal; Zymogen; EGF-like domain; Repeat; 3D-structure; Polymorphism.
FT   SIGNAL       1    31       POTENTIAL.
FT   PROPEP      32    40
FT   CHAIN       41   179       FACTOR X LIGHT CHAIN.
FT   CHAIN      183   488       FACTOR X HEAVY CHAIN.
FT   PROPEP     183   234       ACTIVATION PEPTIDE.
FT   CHAIN      235   488       ACTIVATED FACTOR XA, HEAVY CHAIN.
FT   DOMAIN      86   122       EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     125   165       EGF-LIKE 2.
FT   DOMAIN     235   488       SERINE PROTEASE.
FT   MOD_RES     46    46       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     47    47       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     54    54       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     56    56       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     59    59       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     60    60       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     65    65       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     66    66       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     69    69       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     72    72       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     79    79       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES    103   103       HYDROXYLATION.
FT   CARBOHYD   199   199       O-LINKED (GALNAC. . .).
FT   CARBOHYD   211   211       O-LINKED (GALNAC. . .).
FT   CARBOHYD   221   221       N-LINKED (GLCNAC. . .).
FT                              /FTId=CAR_000012.
FT   CARBOHYD   231   231       N-LINKED (GLCNAC. . .).
FT                              /FTId=CAR_000013.
FT   ACT_SITE   276   276       CHARGE RELAY SYSTEM.
FT   ACT_SITE   322   322       CHARGE RELAY SYSTEM.
FT   ACT_SITE   419   419       CHARGE RELAY SYSTEM.
FT   DISULFID    90   101
FT   DISULFID    95   110
FT   DISULFID   112   121
FT   DISULFID   129   140
FT   DISULFID   136   149
FT   DISULFID   151   164
FT   DISULFID   172   342       INTERCHAIN.
FT   DISULFID   241   246
FT   DISULFID   261   277
FT   DISULFID   390   404
FT   DISULFID   415   443
FT   VARIANT      7     7       L -> I (in dbSNP:5963).
FT                              /FTId=VAR_014162.
FT   VARIANT     30    30       Q -> H (in dbSNP:5961).
FT                              /FTId=VAR_014163.
FT   CONFLICT   285   288       KVRV -> E (IN REF. 6 AND 8).
FT   CONFLICT   442   442       G -> S (IN REF. 5).
FT   HELIX      133   135
FT   STRAND     139   143
FT   TURN       144   145
FT   STRAND     146   150
FT   TURN       153   154
FT   STRAND     155   157
FT   TURN       159   160
FT   STRAND     164   166
```

```
FT   TURN        172    173
FT   STRAND      236    236
FT   STRAND      239    240
FT   TURN        243    244
FT   TURN        247    248
FT   STRAND      249    254
FT   TURN        255    256
FT   STRAND      259    265
FT   STRAND      270    273
FT   HELIX       275    280
FT   STRAND      284    288
FT   STRAND      292    292
FT   TURN        293    294
FT   STRAND      301    311
FT   TURN        312    313
FT   TURN        316    318
FT   STRAND      324    328
FT   STRAND      335    335
FT   TURN        336    337
FT   STRAND      338    338
FT   STRAND      342    342
FT   HELIX       346    352
FT   TURN        353    355
FT   STRAND      358    363
FT   STRAND      376    376
FT   STRAND      378    384
FT   HELIX       387    393
FT   TURN        400    401
FT   STRAND      402    402
FT   STRAND      405    405
FT   STRAND      413    413
FT   TURN        416    420
FT   STRAND      422    426
FT   STRAND      431    440
FT   TURN        446    447
FT   STRAND      450    454
FT   HELIX       456    458
FT   HELIX       459    465
FT   TURN        466    467
FT   STRAND      473    473
SQ   SEQUENCE   488 AA;  54731 MW;  F81D5746AF4797AF CRC64;

FA10_HUMAN  Length: 488  October 5, 2004 14:21  Type: P  Check: 2459  ..

       1  MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

      51  GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

     101  CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS

     151  CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW

     201  KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA

     251  LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD RNTEQEEGGE

     301  AVHEVEVVIK HNRFTKETYD FDIAVLRLKT PITFRMNVAP ACLPERDWAE

     351  STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDRNSC KLSSSFIITQ

     401  NMFCAGYDTK QEDACQGDSG GPHVTRFKDT YFVTGIVSWG EGCARKGKYG

     451  IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
ID   P100_HSV6U     STANDARD;      PRT;   870 AA.
AC   Q00701;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Large structural phosphoprotein (PP100) (P100) (Major antigenic
DE   structural protein).
GN   U11 OR P1LF1.
OS   Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Roseolovirus.
OX   NCBI_TaxID=10370;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92260671; PubMed=1374813;
RA   Neipel F., Ellinger K., Fleckenstein B.;
RT   "Gene for the major antigenic structural protein (p100) of human
RT   herpesvirus 6.";
RL   J. Virol. 66:3918-3924(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94118404; PubMed=8289364;
RA   Nicholas J., Martin M.;
RT   "Nucleotide sequence analysis of a 38.5-kilobase-pair region of the
RT   genome of human herpesvirus 6 encoding human cytomegalovirus
RT   immediate-early gene homologs and transactivating functions.";
RL   J. Virol. 68:597-610(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95266321; PubMed=7747482;
RA   Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RA   Martin M.E., Efstathiou S., Craxton M., Macaulay H.A.;
RT   "The DNA sequence of human herpesvirus-6: structure, coding content,
RT   and genome evolution.";
RL   Virology 209:29-51(1995).
CC   -!- SIMILARITY: TO THE LARGE STRUCTURAL PHOSPHOPROTEINS OF HSV-7 AND
CC       HCMV UL32.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M87287; AAA46012.1; -.
DR   EMBL; L25528; AAA16716.1; ALT_INIT.
DR   EMBL; X83413; CAA58438.1; -.
KW   Matrix protein; Phosphorylation.
SQ   SEQUENCE   870 AA;  97071 MW;  F25954DEA19BF824 CRC64;

P100_HSV6U  Length: 870  October 5, 2004 14:22  Type: P  Check: 500  ..

       1  MDLQRHPIPF AWLDRDKVER LTDFLSNLER LDNVDLREHP HVTNSCVVRE

      51  GDDVDDLKTL YNLLVLWLMY HYVLSKRKPD YNAIWQDITK LQSVVNEYLN

     101  SKGLNKGIFE NMFTNKEKFE SQFSDINRAL LRLGNFIKWG SNVAIDTPYV

     151  NLTAEDSSEI ENNLQDAEKN MLWYTVYNIN DPWDENGYLI TSINKLIYLG

     201  KLFLALTQSW SKLEKVAMSQ IVITQNHLSG HLRRHDNFNI VYSHRVLQTP

     251  LTGQRVESFL KIITSDYDII KSSLESHSAS KAFSMSEIGP NSLMDFVPLR

     301  GDIHSNLTLP SMSIDTKKSS LDPARLKKSN SRSLDSFLRM QRQPKFLELD

     351  SVDNAGEKIL LKEATLGGEN VKATTPASSV SLMSGVESPS SFTSTNLDLP

     401  LSSFTSTNLD LRDKSHGNYK IGPSGILDFN VKFPPNAQLN TNGVDLLQDK

     451  TSIGSPSSGI TDVVNGFANL NLHQNKSNVS PPWSRNTAAN ADFLDPVHRF

     501  VPEQTGTPFV LNNSDVAGSE AKHTTYSTET GVSPRNVFLI KDLRGKDGFR

     551  KQKQSDIPKS LTKERNDKAI MHSREVTGDS GDATETVGAR NSPALRKIKQ

     601  ANDFFAGLNK KNDRDVLRGG KGNSKDLHSG GNAKKKEMSG KFNDDKEMTR

     651  NGQEPSRSLM GDARNAGDEQ YIQAGLGQRV NNLLSQFTNL ISLGEKGIED
```

```
     701  ILQNQRGTEL KLATENKSGR ESEEANVEKI LEVSNPQDMF KNFRLQNDLD

     751  SVQSPFRLPD ADLSRELDSA SFKDALDLKL PGNGEREIDL ALEKVKVGET

     801  ETSDLKVGQD ESFVPAQLMK VETPEEKDDI IEQMVLRIRQ DGETDENTVS

     851  GPGVAESLDI EAKGESAIAS

!!AA_SEQUENCE 1.0
ID   PFDA_METKA      STANDARD;      PRT;   157 AA.
AC   Q8TUY7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Prefoldin alpha subunit (GimC alpha subunit).
GN   PFDA OR MK1614.
OS   Methanopyrus kandleri.
OC   Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC   Methanopyrus.
OX   NCBI_TaxID=2320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AV19 / DSM 6324 / JCM 9639;
RX   MEDLINE=21927647; PubMed=11930014;
RA   Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA   Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA   Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA   Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT   "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT   and monophyly of archaeal methanogens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC   -!- FUNCTION: Molecular chaperone capable of stabilizing a range of
CC       proteins. Seems to fulfil an ATP-independent, HSP70-like function
CC       in archaeal de novo protein folding (By similarity).
CC   -!- SUBUNIT: Heterohexamer of two alpha and four beta subunits (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the prefoldin alpha subunit family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE010451; AAM02827.1; -.
DR   HAMAP; MF_00308; -; 1.
DR   InterPro; IPR004127; DUF232.
DR   Pfam; PF02996; Prefoldin; 1.
KW   Chaperone; Complete proteome.
SQ   SEQUENCE   157 AA;  17446 MW;  B20BDA8CC978DFA1 CRC64;

PFDA_METKA  Length: 157  October 5, 2004 14:22  Type: P  Check: 9337  ..

       1  MAEKKNEQEI QQELQRLIAE INRLQGQMEA INAQIDLIES SISELNRVEE

      51  TLKGVKELEG DEEVLVPVGA QSFVRACVTD TERVIVGIGA GVAVERTIDE

     101  ALESIDDQRQ ELEKARAEAQ QKLQELAQEL QEKQRKAQEL AQQLEGAQRI

     151  AQQSGGG

!!AA_SEQUENCE 1.0
ID   VHS_VZVD        STANDARD;      PRT;   455 AA.
AC   P09275;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Virion host shutoff protein.
GN   17.
OS   Varicella-zoster virus (strain Dumas) (VZV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10338;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86306657; PubMed=3018124;
RA   Davison A.J., Scott J.E.;
RT   "The complete DNA sequence of varicella-zoster virus.";
RL   J. Gen. Virol. 67:1759-1816(1986).
CC   -!- FUNCTION: REQUIRED FOR INHIBITION OF CELLULAR GENE EXPRESSION IN
CC       INFECTED CELLS.
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL41,
CC       EHV-1 19, AND VZV 17.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04370; CAA27900.1; -.
DR   PIR; H27342; WZBE17.
DR   InterPro; IPR000513; Exo_N_I.
DR   InterPro; IPR002927; Virus_HS.
DR   Pfam; PF01550; Virus_HS; 1.
SQ   SEQUENCE   455 AA;  51368 MW;  9D9EE755903DC4C7 CRC64;

 VHS_VZVD  Length: 455  October 5, 2004 14:22  Type: P  Check: 3936  ..

       1  MGLFGLTRFI HEHKLVKPSI ISTPPGVLTP VAVDVWNVMY TLLERLYPVG

      51  KRENLHGPSV TIHCLGVLLR LLTQRSYYPI FVLERCTDGP LSRGAKAIMS

     101  RAMNHDERGT SDLTRVLLSS NTSCSIKYNK TSETYDSVFR NSSTSCIPSE

     151  ENKSQDMFLD GCPRQTDKTI CLRDQNVCSL TSTMPSRGHP NHRLYHKLCA

     201  SLIRWMGYAY VEAVDIEADE ACANLFHTRT VALVYTTDTD LLFMGCDILL

     251  DAIPMFAPVV RCRDLLQYLG ITYPEFLVAF VRCQTDLHTS DNLKSVQQVI

     301  QDTGLKVPHQ MDTSTRSPTY DSWRHGEVFK SLTVATSGKT ENGVSVSKYA

     351  SNRSEVTVDA SWALNLLPPS SSPLDNLERA FVEHIIAVVT PLTRGRLKLM

     401  KRVNIMQNTA DPYMVINTLY HNLKGEKMAR QYARIFKQFI PTPLPLNTVL

     451  TKYWN

!!AA_SEQUENCE 1.0
ID   VPX_HV2D2       STANDARD;      PRT;   111 AA.
AC   P15836;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   VPX protein (X ORF protein) (Viral accessory protein).
GN   VPX.
OS   Human immunodeficiency virus type 2 (isolate D205,7) (HIV-2).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11716;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90081881; PubMed=2594088;
RA   Dietrich U., Adamski M., Kreutz R., Seipp A., Kuehnel H.,
RA   Ruebsamen-Waigmann H.;
RT   "A highly divergent HIV-2-related isolate.";
RL   Nature 342:948-950(1989).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X61240; -; NOT_ANNOTATED_CDS.
DR   PIR; S08438; S08438.
DR   HIV; X16109; VPX$2D205.
DR   InterPro; IPR000012; RetroV_VpR/X.
DR   Pfam; PF00522; VPR; 1.
KW   AIDS.
SQ   SEQUENCE   111 AA;  12856 MW;  D67E9921ED15FEF6 CRC64;

VPX_HV2D2  Length: 111  October 5, 2004 14:22  Type: P  Check: 2803  ..

       1  MDPRERVPPG NSDEETVGEA FAWLERTITE LNRVAVNHLP RELIFQVWQR

      51  SWAYWREEQG MSISYTKYRY LLLMQKAMFV HYTKGCRCLQ EGHGPGGWRS

     101  GPPPPPPPGL A

!!AA_SEQUENCE 1.0
ID   YCAO_ECOLI     STANDARD;      PRT;   586 AA.
AC   P75838; Q9R2W0;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ycaO.
GN   YCAO OR B0905.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
CC   -!- SIMILARITY: BELONGS TO THE UPF0142 FAMILY. STRONG, TO H.INFLUENZAE
CC       HI1265.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000192; AAC73991.1; ALT_INIT.
DR   EMBL; D90728; BAA35640.1; ALT_INIT.
DR   EMBL; D90729; BAA35649.1; ALT_INIT.
DR   EcoGene; EG13699; ycaO.
DR   InterPro; IPR003776; DUF181.
DR   Pfam; PF02624; DUF181; 1.
DR   TIGRFAMs; TIGR00702; TIGR00702; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   586 AA;  65651 MW;  44CD903F9E73F503 CRC64;

YCAO_ECOLI  Length: 586  October 5, 2004 14:22  Type: P  Check: 8500  ..

       1  MTQTFIPGKD AALEDSIARF QQKLSDLGFQ IEEASWLNPV PNVWSVHIRD

      51  KECALCFTNG KGATKKAALA SALGEYFERL STNYFFADFW LGETIANGPF

     101  VHYPNEKWFP LTENDDVPEG LLDDRLRAFY DPENELTGSM LIDLQSGNED

     151  RGICGLPFTR QSDNQTVYIP MNIIGNLYVS NGMSAGNTRN EARVQGLSEV

     201  FERYVKNRII AESISLPEIP ADVLARYPAV VEAIETLEAE GFPIFAYDGS

     251  LGGQYPVICV VLFNPANGTC FASFGAHPDF GVALERTVTE LLQGRGLKDL

     301  DVFTPPTFDD EEVAEHTNLE THFIDSSGLI SWDLFKQDAD YPFVDWNFSG

     351  TTEEEFATLM AIFNKEDKEV YIADYEHLGV YACRIIVPGM SDIYPAEDLW

     401  LANNSMGSHL RETILSLPGS EWEKEDYLNL IEQLDEEGFD DFTRVRELLG

     451  LATGSDNGWY TLRIGELKAM LALAGGDLEQ ALVWTEWTME FNSSVFSPER

     501  ANYYRCLQTL LLLAQEEDRQ PLQYLNAFVR MYGADAVEAA SAAMSGEAAF

     551  YGLQPVDSDL HAFAAHQSLL KAYEKLQRAK AAFWAK

!!AA_SEQUENCE 1.0
ID   YZN4_CAEEL     STANDARD;      PRT;   692 AA.
AC   P54245;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical acetylcholine receptor like protein F18G5.4 in
DE   chromosome X.
GN   F18G5.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Favello A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Possible acetylcholine receptor.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U39855; AAA81080.1; -.
DR   PIR; A89606; A89606.
DR   WormPep; F18G5.4; CE04412.
DR   InterPro; IPR006029; Neu_channel_memb.
DR   InterPro; IPR006202; Neur_chan_LBD.
DR   InterPro; IPR006201; Neur_channel.
DR   Pfam; PF02931; Neur_chan_LBD; 1.
DR   Pfam; PF02932; Neur_chan_memb; 1.
DR   PRINTS; PR00252; NRIONCHANNEL.
DR   PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
KW   Hypothetical protein; Receptor; Postsynaptic membrane; Ionic channel;
KW   Glycoprotein; Transmembrane.
```

```
FT   TRANSMEM     39     59       POTENTIAL.
FT   TRANSMEM    402    422       POTENTIAL.
FT   TRANSMEM    432    452       POTENTIAL.
FT   TRANSMEM    460    481       POTENTIAL.
FT   DISULFID    315    329       BY SIMILARITY.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    359    359       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    550    550       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    593    593       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   692 AA;  79291 MW;  3510664F780FAD27 CRC64;

YZN4_CAEEL  Length: 692  October 5, 2004 14:23  Type: P  Check: 9797  ..

       1  MKHVASLAHC FFHLGRRCLK LRPHKLEEPR HRSKLVRRAL LPLLLHSFAA

      51  FFFLFCAPFH TITFALHITY VRREDLLLLL PLCICCAPAH HPRSFSNWLL

     101  KQSVQSTGPQ FPPKFQKPHN EENTIGTITK FAPSVQEQHS SAVIPMPHFD

     151  QNRLEQALRI KGSIDGTEEA LYRSLLDHTV YEKDVRPCIH HSQPTNVTFG

     201  FLLNQIVEMD ERNQALTTRS WLNINWMDPR LSWNESLWSE IKAIYIPHAR

     251  IWKPDIILVN NRQTKNSDAC LHAYISASEP AAIREYYASL VSTDVMVTSD

     301  GNVTWLFSAL FRSSCPIRVR YYPFDDQQCD LKFASWSHDI TEINLGLNTD

     351  KGDLSSYMNN SEFDLVDMTA VREVVTFPSD TNSDWPIIVI RIHMHRRPLF

     401  YVFNHIVPCV LISSMAVLGF LMPPETGEKI NMIITTLLSM GVYLQSITES

     451  IPPTSEGVPL IGMYYVSSLL MVCLATCVNV ITLNMHRNGA ANQGRHVPAW

     501  MQKWILGYLA TFMRMSIREP DSIALLKASQ SKKSTIRRSS ILRDLKRVKN

     551  MSNVRAKSKE QNANRECECM DPLVHIYAES IMSCLAADTK PMNGSTIRED

     601  FASESTFLGR VVSDGIMPRI SASSNSVLTE FETRFRRILK RVYRSLQQHE

     651  IREEILDERS RIQCSGNNLH LSLIDFYYVF FALQHCSQSS AS
```

citations for A-Geneseq hits

!!AA_SEQUENCE 1.0
ID **AAB10470** standard; protein; 543 AA.
XX
AC AAB10470;
XX
DT 11-DEC-2000 (first entry)
XX
DE Shewanella putrefaciens PKS protein ORF9.
XX
KW PKS pathway; polyunsaturated long chain fatty acid; plant; transgenic;
KW polyketide-like synthesis; PUFA; dietary supplement; intravenous feeding;
KW malnutrition; cooking oil; cooking fat; margarine;
KW docosahexenoic acid production; eicosapentenoic acid production.
XX
OS Shewanella putrefaciens.
XX
PN WO200042195-A2.
XX
PD 20-JUL-2000.
XX
PF 14-JAN-2000; 2000WO-US000956.
XX
PR 14-JAN-1999; 99US-00231899.
XX
PA (CALJ) CALGENE LLC.
XX
PI Facciotti D, Metz JG, Lassner M;
XX
DR WPI; 2000-476063/41.
XX
PT New DNA sequences encoding for polyketide (PK)-like synthesis pathway
PT genes from Shewanella, Vibrio and Schizochtrium, useful for creating
PT transgenic plants that express poly-unsaturated long chain fatty acids.
XX
PS Example 1; Fig 4J; 302pp; English.
XX
CC This invention describes novel DNA sequences encoding for polyketide (PK)
CC -like synthesis (PKS-like) pathway genes from Shewanella, Vibrio and
CC Schizochtrium. The nucleic acids are useful for isolating related
CC molecules or in methods to detect organisms expressing the PKS-like
CC genes. They are also useful for creating transgenic plants that express
CC poly-unsaturated long chain fatty acids. The poly-unsaturated long chain
CC fatty acids produced recombinantly are useful as dietary supplements for
CC patients undergoing intravenous feeding or for preventing or treating
CC malnutrition. The poly-unsaturated long chain fatty acids can also be
CC incorporated into cooking oils, fats or margarine formulated so that in
CC normal use the recipient receives a desired amount of poly-unsaturated
CC long chain fatty acids. The nucleic acids are also useful in large scale
CC production of docosahexenoic acid and eicosapentenoic acid, and for the
CC modification of the fatty acid profile of host cells and edible plant
CC tissues and/or plant parts. Transgenic production of polyunsaturated
CC fatty acids in particular host cells allows quicker purification from
CC natural sources such as fish or plants. This sequence represents the
CC Shewanella putrefaciens PKS protein cluster ORF9 which is described in
CC the method of the invention
XX
SQ Sequence 543 AA;

AAB10470 Length: 543 October 5, 2004 14:15 Type: P Check: 3609 ..

1 MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRAC YVVNHADHGF

51 GIAQTADIVT EQAANSTDLP VSAFTPALGT ESLGDNNFRR VHGVKYAYYA

101 GAMANGISSE ELVIALGQAG ILCGSFGAAG LIPSRVEAAI NRIQAALPNG

151 PYMFNLIHSP SEPALERGSV ELFLKHKVRT VEASAFLGLT PQIVYYRAAG

201 LSRDAQGKVV VGNKVIAKVS RTEVAEKFMM PAPAKMLQKL VDDGSITAEQ

251 MELAQLVPMA DDITAEADSG GHTDNRPLVT LLPTILALKE EIQAKYQYDT

301 PIRVGCGGGV GTPDAALATF NMGAAYIVTG SINQACVEAG ASDHTRKLLA

351 TTEMADVTMA PAADMFEMGV KLQVVKRGTL FPMRANKLYE IYTRYDSIEA

401 IPLDEREKLE KQVFRSSLDE IWAGTVAHFN ERDPKQIERA EGNPKRKMAL

451 IFRWYLGLSS RWSNSGEVGR EMDYQIWAGP ALGAFNQWAK GSYLDNYQDR

501 NAVDLAKHLM YGAAYLNRIN SLTAQGVKVP AQLLRWKPNQ RMA

!!AA_SEQUENCE 1.0
ID **AAB70411** standard; protein; 488 AA.
XX
AC AAB70411;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human factor X protein sequence SEQ ID NO:2.
XX
KW Human; factor X; mutant; haemostatic; gene therapy; haemophilia;
KW blood coagulation disorder; haemophiliac.
XX
OS Homo sapiens.
XX
PN WO200110896-A2.
XX
PD 15-FEB-2001.
XX
PF 07-AUG-2000; 2000WO-EP007631.
XX
PR 10-AUG-1999; 99AT-00001377.
XX
PA (BAXT) BAXTER AG.
XX
PI Himmelspach M, Schlokat U;
XX
DR WPI; 2001-191516/19.
DR N-PSDB; AAF59409.
XX
PT Novel factor X analog useful for producing drug which is useful for
PT treatment of blood coagulation disorders, such as hemophilia, contains
PT modification between amino acids Glu226 and Ile235.
XX
PS Disclosure; Fig 1; 50pp; English.
XX
CC The present invention describes a factor X analogue (I) which contains a
CC modification between Glu226 and Ile235, relative to the 488 residue amino
CC acid sequence given in AAB70411. (I) has haemostatic activity and can be
CC used in gene therapy. (I) encoding polynucleotide (II) can be used to
CC produce a drug, which is useful for treatment of patients with blood
CC coagulation disorders, such as patients suffering from haemophilia, or
CC haemophiliacs with inhibitory antibodies. Preparations containing a
CC polypeptide with factor X/Xa activity are more readily activated by
CC factor X1a or its derivative, which has high stability, without having to
CC use one of the proteases used in prior art to activate the natural factor
CC X, particularly one of animal origins, such as Russell's viper venom
CC (RVV) or trypsin. The present sequence represents human factor X, which
CC is given in the exemplification of the present invention
XX
SQ Sequence 488 AA;

AAB70411 Length: 488 October 5, 2004 14:16 Type: P Check: 2975 ..

1 MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

51 GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

101 CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS

151 CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW

201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA

251 LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD RNTEQEEGGE

```
     301  AVHEVEVVIK HNRFTKETYD FDIAVLRLKT PITFRMNVAP ACLPERDWAE

     351  STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDRNSC KLSSSFIITQ

     401  NMFCAGYDTK QEDACQGDSG GPHVTRFKDT YFVTGIVSWG ESCARKGKYG

     451  IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
ID   AAE12822 standard; protein; 635 AA.
XX
AC   AAE12822;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Caenorhabditis elegans 5-HT3 receptor protein, F18.
XX
KW   5-hydroxytryptamine 3; 5-HT3 receptor; nematocidal; insecticidal;
KW   pesticidal; antihelminthic.
XX
OS   Caenorhabditis elegans.
XX
PN   WO200161000-A1.
XX
PD   23-AUG-2001.
XX
PF   15-FEB-2001; 2001WO-AU000150.
XX
PR   15-FEB-2000; 2000AU-00005634.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Trowell SC,  Dumancic MM,  Liao C,  East PD;
XX
DR   WPI; 2001-648195/74.
DR   N-PSDB; AAD20960.
XX
PT   New invertebrate 5-hydroxytryptamine (5-HT3) receptor, especially from
PT   Caenorhabditis elegans, is useful for identifying and/or assessing
PT   candidate compounds for nematicidal, insecticidal or pesticidal use.
XX
PS   Claim 18; Page 64-66; 74pp; English.
XX
CC   The patent discloses novel invertebrate 5-hydroxytryptamine 3 (5-HT3)
CC   receptor proteins, especially from Caenorhabditis elegans and their
CC   corresponding polynucleotides. The 5-HT3 receptors are useful in assays
CC   for identifying and/or assessing candidate compounds for nematocidal,
CC   insecticidal or pesticidal use. The present invention also relates to
CC   methods for killing a helminth (particularly a nematode) or an insect,
CC   particularly a sucking insect or other insect with a muscular pump- based
CC   feeding mechanism by exposing them to an effective amount of a compound
CC   which alters their 5-HT3 receptor activity. The present sequence is 5-HT3
CC   receptor protein, F18 from Caenorhabditis elegans
XX
SQ   Sequence 635 AA;

 AAE12822  Length: 635  October 5, 2004 14:16  Type: P  Check: 9069  ..

       1  MIICYSCLTV SILLTIKFVP CRFAGIEHQN TKSRVHFSLL DSRQENDTNH

      51  FEIAEAKFQK PHNEENTIGT ITKFAPSVQE QHSSAVIPMP HFDQNRLEQA

     101  LRIKGSIDGT EEALYRSLLD HTVYEKDVRP CIHHSQPTNV TFGFLLNQIV

     151  EMDERNQALT TRSWLNINWM DPRLSWNESL WSEIKAIYIP HARIWKPDII

     201  LVNNAIREYY ASLVSTDVMV TSDGNVTWLF SALFRSSCPI RVRYYPFDDQ

     251  QCDLKFASWS HDITEINLGL NTDKGDLSSY MNNSEFDLVD MTAVREVVTF

     301  PSDTNSDWPI IVIRIHMHRR PLFYVFNHIV PCVLISSMAV LGFLMPPETG

     351  EKINMIITTL LSMGVYLQSI TESIPPTSEG VPLIGMYYVS SLLMVCLATC

     401  VNVITLNMHR NGAANQGRHV PAWMQKWILG YLATFMRMSI REPDSIALLK

     451  ASQSKKSTIR RSSILRDLKR VKNMSNVRAK SKEQNANREC ECMDPLVHIY

     501  AESIMSCLAA DTKPMNGSTI REDFASESTF LGRVVSDGIM PRISASSNSV

     551  LTEFETRFRR ILKRVYRSLQ QHEIREEILD ERSRIQWQWQ QLASVVDRLL

     601  LCLFCTATLF TIICLLIVPV AYRDNDSMLS FLNFF

!!AA_SEQUENCE 1.0
ID   AAG52216 standard; protein; 274 AA.
XX
AC   AAG52216;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 66352.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
```

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162143P.
PR 29-OCT-1999; 99US-0162228P.
PR 01-NOV-1999; 99US-0162891P.
PR 01-NOV-1999; 99US-0162894P.
PR 01-NOV-1999; 99US-0162895P.
PR 02-NOV-1999; 99US-0163091P.

PR 02-NOV-1999; 99US-0163092P.
PR 02-NOV-1999; 99US-0163093P.
PR 03-NOV-1999; 99US-0163248P.
PR 03-NOV-1999; 99US-0163249P.
PR 03-NOV-1999; 99US-0163281P.
PR 04-NOV-1999; 99US-0163379P.
PR 04-NOV-1999; 99US-0163380P.
PR 04-NOV-1999; 99US-0163381P.
PR 08-NOV-1999; 99US-0164146P.
PR 08-NOV-1999; 99US-0164150P.
PR 08-NOV-1999; 99US-0164151P.
PR 09-NOV-1999; 99US-0164259P.
PR 09-NOV-1999; 99US-0164260P.
PR 10-NOV-1999; 99US-0164317P.
PR 10-NOV-1999; 99US-0164318P.
PR 10-NOV-1999; 99US-0164319P.
PR 10-NOV-1999; 99US-0164321P.
PR 10-NOV-1999; 99US-0164544P.
PR 10-NOV-1999; 99US-0164545P.
PR 10-NOV-1999; 99US-0164548P.
PR 12-NOV-1999; 99US-0164870P.
PR 12-NOV-1999; 99US-0164871P.
PR 12-NOV-1999; 99US-0164959P.
PR 12-NOV-1999; 99US-0164960P.
PR 12-NOV-1999; 99US-0164961P.
PR 12-NOV-1999; 99US-0164962P.
PR 15-NOV-1999; 99US-0164926P.
PR 15-NOV-1999; 99US-0164927P.
PR 15-NOV-1999; 99US-0164929P.
PR 16-NOV-1999; 99US-0165661P.
PR 16-NOV-1999; 99US-0165669P.
PR 16-NOV-1999; 99US-0165671P.
PR 17-NOV-1999; 99US-0165911P.
PR 17-NOV-1999; 99US-0165918P.
PR 17-NOV-1999; 99US-0165919P.
PR 18-NOV-1999; 99US-0166157P.
PR 18-NOV-1999; 99US-0166158P.
PR 18-NOV-1999; 99US-0166173P.
PR 19-NOV-1999; 99US-0166411P.
PR 19-NOV-1999; 99US-0166412P.
PR 19-NOV-1999; 99US-0166419P.
PR 22-NOV-1999; 99US-0166733P.
PR 22-NOV-1999; 99US-0166750P.
PR 23-NOV-1999; 99US-0167362P.
PR 24-NOV-1999; 99US-0167233P.
PR 24-NOV-1999; 99US-0167234P.
PR 24-NOV-1999; 99US-0167235P.
PR 24-NOV-1999; 99US-0167382P.
PR 30-NOV-1999; 99US-0167902P.
PR 30-NOV-1999; 99US-0167904P.
PR 30-NOV-1999; 99US-0167908P.
PR 01-DEC-1999; 99US-0168231P.
PR 01-DEC-1999; 99US-0168232P.
PR 01-DEC-1999; 99US-0168233P.
PR 02-DEC-1999; 99US-0168546P.
PR 02-DEC-1999; 99US-0168548P.
PR 02-DEC-1999; 99US-0168549P.
PR 03-DEC-1999; 99US-0168673P.
PR 03-DEC-1999; 99US-0168674P.
PR 03-DEC-1999; 99US-0168675P.
PR 07-DEC-1999; 99US-0169278P.
PR 07-DEC-1999; 99US-0169298P.
PR 07-DEC-1999; 99US-0169302P.
PR 08-DEC-1999; 99US-0169691P.
PR 08-DEC-1999; 99US-0169692P.
PR 16-DEC-1999; 99US-0171098P.
PR 16-DEC-1999; 99US-0171107P.
PR 16-DEC-1999; 99US-0171114P.
PR 19-JAN-2000; 2000US-0176866P.
PR 19-JAN-2000; 2000US-0176867P.
PR 19-JAN-2000; 2000US-0176910P.
PR 26-JAN-2000; 2000US-0178166P.
PR 27-JAN-2000; 2000US-0177666P.
PR 27-JAN-2000; 2000US-0178544P.
PR 27-JAN-2000; 2000US-0178545P.
PR 27-JAN-2000; 2000US-0178546P.
PR 27-JAN-2000; 2000US-0178547P.
PR 28-JAN-2000; 2000US-0178754P.
PR 28-JAN-2000; 2000US-0178755P.
PR 01-FEB-2000; 2000US-0179388P.
PR 01-FEB-2000; 2000US-0179395P.
PR 03-FEB-2000; 2000US-0180039P.
PR 03-FEB-2000; 2000US-0180139P.
PR 04-FEB-2000; 2000US-0180206P.
PR 04-FEB-2000; 2000US-0180207P.
PR 07-FEB-2000; 2000US-0180695P.
PR 07-FEB-2000; 2000US-0180696P.
PR 09-FEB-2000; 2000US-0181214P.
PR 09-FEB-2000; 2000US-0181228P.
PR 10-FEB-2000; 2000US-0181476P.
PR 10-FEB-2000; 2000US-0181551P.
PR 15-FEB-2000; 2000US-0182477P.
PR 15-FEB-2000; 2000US-0182478P.
PR 15-FEB-2000; 2000US-0182512P.
PR 15-FEB-2000; 2000US-0182516P.
PR 17-FEB-2000; 2000US-0183165P.
PR 17-FEB-2000; 2000US-0183166P.
XX
PA (CERE-) CERES INC.
XX
PI Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L, Dumas J;
XX
DR WPI; 2000-507395/46.
DR N-PSDB; AAC50846.
XX
PT New sequence determined DNA fragments (SDFs) from different plant
PT species, e.g. corn, rice or Arabidopsis thaliana, useful as promoters,
PT protein coding sequences, untranslated regions, or as 3' termination
PT sequences.
XX
PS Claim 19; SEQ ID NO 66352; 344pp + Sequence Listing; English.
XX
CC The present sequence is a putative protein fragment from Arabidopsis
CC thaliana. Its coding sequence was isolated by carrying out RT-PCR on all
CC of the mRNA obtained from the plant, and creating a cDNA library which
CC could then be sequenced, allowing the putative protein sequence(s) to be
CC obtained. This sequence may be useful for protein identification and for
CC aiding in the elucidation of signal transduction and metabolic pathways.
CC Its coding sequence has a use in the control of gene expression as a
CC promoter, coding sequence, 3'UTR or termination sequence, for controlling
CC the behaviour of a gene within the chromosome, as a tool for use in
CC genetic mapping, including a use in hybridisation assays, for recognition
CC or isolation of similar DNA fragments, or for the identification of a
CC particular organism
XX
SQ Sequence 274 AA;

AAG52216 Length: 274 October 5, 2004 14:15 Type: P Check: 7182 ..

1 MGDFSIQISS KLINQLAEGN DQPKRRAKKT KPKVSPQSKQ KTNQDEEKKP

51 NPVAELPMQP PFFFPIPPQG AASTELESIK SVVKESEKVL EKLELQEKNI

101 VREVTERAKD LREKEFKIPE PKPMPCSSDH EAWMKCYKEN IGSPLKCSGF

151 VKSFQDCARR SRQQVNPEEN SNTLNRVNLG EQIFLSIFNV MTRMMLGAIV

201 EEEERTILGN ELKKLILLPQ ISKEAQKYSY VPHRNSKTSV VAGYTVLKDL

251 KIFINVWAIN RDTKNWEEEG SKRV

!!AA_SEQUENCE 1.0
ID **AAG52217** standard; protein; 217 AA.
XX
AC AAG52217;

XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66353.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.

PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162143P.
PR 29-OCT-1999; 99US-0162228P.
PR 01-NOV-1999; 99US-0162891P.
PR 01-NOV-1999; 99US-0162894P.
PR 01-NOV-1999; 99US-0162895P.
PR 02-NOV-1999; 99US-0163091P.
PR 02-NOV-1999; 99US-0163092P.
PR 02-NOV-1999; 99US-0163093P.
PR 03-NOV-1999; 99US-0163248P.
PR 03-NOV-1999; 99US-0163249P.
PR 03-NOV-1999; 99US-0163281P.
PR 04-NOV-1999; 99US-0163379P.
PR 04-NOV-1999; 99US-0163380P.
PR 04-NOV-1999; 99US-0163381P.
PR 08-NOV-1999; 99US-0164146P.
PR 08-NOV-1999; 99US-0164150P.
PR 08-NOV-1999; 99US-0164151P.
PR 09-NOV-1999; 99US-0164259P.
PR 09-NOV-1999; 99US-0164260P.
PR 10-NOV-1999; 99US-0164317P.
PR 10-NOV-1999; 99US-0164318P.
PR 10-NOV-1999; 99US-0164319P.
PR 10-NOV-1999; 99US-0164321P.
PR 10-NOV-1999; 99US-0164544P.
PR 10-NOV-1999; 99US-0164545P.
PR 10-NOV-1999; 99US-0164548P.
PR 12-NOV-1999; 99US-0164870P.
PR 12-NOV-1999; 99US-0164871P.
PR 12-NOV-1999; 99US-0164959P.
PR 12-NOV-1999; 99US-0164960P.
PR 12-NOV-1999; 99US-0164961P.
PR 12-NOV-1999; 99US-0164962P.
PR 15-NOV-1999; 99US-0164926P.
PR 15-NOV-1999; 99US-0164927P.
PR 15-NOV-1999; 99US-0164929P.
PR 16-NOV-1999; 99US-0165661P.
PR 16-NOV-1999; 99US-0165669P.
PR 16-NOV-1999; 99US-0165671P.
PR 17-NOV-1999; 99US-0165911P.
PR 17-NOV-1999; 99US-0165918P.
PR 17-NOV-1999; 99US-0165919P.
PR 18-NOV-1999; 99US-0166157P.
PR 18-NOV-1999; 99US-0166158P.
PR 18-NOV-1999; 99US-0166173P.
PR 19-NOV-1999; 99US-0166411P.
PR 19-NOV-1999; 99US-0166412P.
PR 19-NOV-1999; 99US-0166419P.
PR 22-NOV-1999; 99US-0166733P.
PR 22-NOV-1999; 99US-0166750P.
PR 23-NOV-1999; 99US-0167362P.
PR 24-NOV-1999; 99US-0167233P.
PR 24-NOV-1999; 99US-0167234P.
PR 24-NOV-1999; 99US-0167235P.
PR 24-NOV-1999; 99US-0167382P.
PR 30-NOV-1999; 99US-0167902P.
PR 30-NOV-1999; 99US-0167904P.
PR 30-NOV-1999; 99US-0167908P.
PR 01-DEC-1999; 99US-0168231P.
PR 01-DEC-1999; 99US-0168232P.
PR 01-DEC-1999; 99US-0168233P.
PR 02-DEC-1999; 99US-0168546P.
PR 02-DEC-1999; 99US-0168548P.
PR 02-DEC-1999; 99US-0168549P.
PR 03-DEC-1999; 99US-0168673P.
PR 03-DEC-1999; 99US-0168674P.
PR 03-DEC-1999; 99US-0168675P.
PR 07-DEC-1999; 99US-0169278P.
PR 07-DEC-1999; 99US-0169298P.
PR 07-DEC-1999; 99US-0169302P.
PR 08-DEC-1999; 99US-0169691P.
PR 08-DEC-1999; 99US-0169692P.
PR 16-DEC-1999; 99US-0171098P.
PR 16-DEC-1999; 99US-0171107P.
PR 16-DEC-1999; 99US-0171114P.
PR 19-JAN-2000; 2000US-0176866P.
PR 19-JAN-2000; 2000US-0176867P.
PR 19-JAN-2000; 2000US-0176910P.
PR 26-JAN-2000; 2000US-0178166P.
PR 27-JAN-2000; 2000US-0177666P.
PR 27-JAN-2000; 2000US-0178544P.
PR 27-JAN-2000; 2000US-0178545P.
PR 27-JAN-2000; 2000US-0178546P.
PR 27-JAN-2000; 2000US-0178547P.
PR 28-JAN-2000; 2000US-0178754P.
PR 28-JAN-2000; 2000US-0178755P.
PR 01-FEB-2000; 2000US-0179388P.
PR 01-FEB-2000; 2000US-0179395P.
PR 03-FEB-2000; 2000US-0180039P.
PR 03-FEB-2000; 2000US-0180139P.
PR 04-FEB-2000; 2000US-0180206P.
PR 04-FEB-2000; 2000US-0180207P.
PR 07-FEB-2000; 2000US-0180695P.
PR 07-FEB-2000; 2000US-0180696P.
PR 09-FEB-2000; 2000US-0181214P.
PR 09-FEB-2000; 2000US-0181228P.

PR 10-FEB-2000; 2000US-0181476P.
PR 10-FEB-2000; 2000US-0181551P.
PR 15-FEB-2000; 2000US-0182477P.
PR 15-FEB-2000; 2000US-0182478P.
PR 15-FEB-2000; 2000US-0182512P.
PR 15-FEB-2000; 2000US-0182516P.
PR 17-FEB-2000; 2000US-0183165P.
PR 17-FEB-2000; 2000US-0183166P.
XX
PA (CERE-) CERES INC.
XX
PI Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L, Dumas J;
XX
DR WPI; 2000-507395/46.
DR N-PSDB; AAC50846.
XX
PT New sequence determined DNA fragments (SDFs) from different plant
PT species, e.g. corn, rice or Arabidopsis thaliana, useful as promoters,
PT protein coding sequences, untranslated regions, or as 3' termination
PT sequences.
XX
PS Claim 19; SEQ ID NO 66353; 344pp + Sequence Listing; English.
XX
CC The present sequence is a putative protein fragment from Arabidopsis
CC thaliana. Its coding sequence was isolated by carrying out RT-PCR on all
CC of the mRNA obtained from the plant, and creating a cDNA library which
CC could then be sequenced, allowing the putative protein sequence(s) to be
CC obtained. This sequence may be useful for protein identification and for
CC aiding in the elucidation of signal transduction and metabolic pathways.
CC Its coding sequence has a use in the control of gene expression as a
CC promoter, coding sequence, 3'UTR or termination sequence, for controlling
CC the behaviour of a gene within the chromosome, as a tool for use in
CC genetic mapping, including a use in hybridisation assays, for recognition
CC or isolation of similar DNA fragments, or for the identification of a
CC particular organism
XX
SQ Sequence 217 AA;

AAG52217 Length: 217 October 5, 2004 14:15 Type: P Check: 376 ..

1 MQPPFFFPIP PQGAASTELE SIKSVVKESE KVLEKLELQE KNIVREVTER

51 AKDLREKEFK IPEPKPMPCS SDHEAWMKCY KENIGSPLKC SGFVKSFQDC

101 ARRSRQQVNP EENSNTLNRV NLGEQIFLSI FNVMTRMMLG AIVEEEERTI

151 LGNELKKLIL LFQISKEAQK YSYVPHRNSK TSVVAGYTVL KDLKIFINVW

201 AINRDTKNWE EEGSKRV

!!AA_SEQUENCE 1.0
ID **AAG52218** standard; protein; 151 AA.
XX
AC AAG52218;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66354.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162143P.
PR 29-OCT-1999; 99US-0162228P.
PR 01-NOV-1999; 99US-0162891P.
PR 01-NOV-1999; 99US-0162894P.
PR 01-NOV-1999; 99US-0162895P.
PR 02-NOV-1999; 99US-0163091P.
PR 02-NOV-1999; 99US-0163092P.
PR 02-NOV-1999; 99US-0163093P.
PR 03-NOV-1999; 99US-0163248P.
PR 03-NOV-1999; 99US-0163249P.
PR 03-NOV-1999; 99US-0163281P.
PR 04-NOV-1999; 99US-0163379P.
PR 04-NOV-1999; 99US-0163380P.
PR 04-NOV-1999; 99US-0163381P.
PR 08-NOV-1999; 99US-0164146P.
PR 08-NOV-1999; 99US-0164150P.
PR 08-NOV-1999; 99US-0164151P.
PR 09-NOV-1999; 99US-0164259P.
PR 09-NOV-1999; 99US-0164260P.
PR 10-NOV-1999; 99US-0164317P.
PR 10-NOV-1999; 99US-0164318P.
PR 10-NOV-1999; 99US-0164319P.
PR 10-NOV-1999; 99US-0164321P.
PR 10-NOV-1999; 99US-0164544P.
PR 10-NOV-1999; 99US-0164545P.
PR 10-NOV-1999; 99US-0164548P.
PR 12-NOV-1999; 99US-0164870P.
PR 12-NOV-1999; 99US-0164871P.
PR 12-NOV-1999; 99US-0164959P.
PR 12-NOV-1999; 99US-0164960P.
PR 12-NOV-1999; 99US-0164961P.
PR 12-NOV-1999; 99US-0164962P.
PR 15-NOV-1999; 99US-0164926P.
PR 15-NOV-1999; 99US-0164927P.
PR 15-NOV-1999; 99US-0164929P.
PR 16-NOV-1999; 99US-0165661P.
PR 16-NOV-1999; 99US-0165669P.
PR 16-NOV-1999; 99US-0165671P.
PR 17-NOV-1999; 99US-0165911P.
PR 17-NOV-1999; 99US-0165918P.

PR 17-NOV-1999; 99US-0165919P.
PR 18-NOV-1999; 99US-0166157P.
PR 18-NOV-1999; 99US-0166158P.
PR 18-NOV-1999; 99US-0166173P.
PR 19-NOV-1999; 99US-0166411P.
PR 19-NOV-1999; 99US-0166412P.
PR 19-NOV-1999; 99US-0166419P.
PR 22-NOV-1999; 99US-0166733P.
PR 22-NOV-1999; 99US-0166750P.
PR 23-NOV-1999; 99US-0167362P.
PR 24-NOV-1999; 99US-0167233P.
PR 24-NOV-1999; 99US-0167234P.
PR 24-NOV-1999; 99US-0167235P.
PR 24-NOV-1999; 99US-0167382P.
PR 30-NOV-1999; 99US-0167902P.
PR 30-NOV-1999; 99US-0167904P.
PR 30-NOV-1999; 99US-0167908P.
PR 01-DEC-1999; 99US-0168231P.
PR 01-DEC-1999; 99US-0168232P.
PR 01-DEC-1999; 99US-0168233P.
PR 02-DEC-1999; 99US-0168546P.
PR 02-DEC-1999; 99US-0168548P.
PR 02-DEC-1999; 99US-0168549P.
PR 03-DEC-1999; 99US-0168673P.
PR 03-DEC-1999; 99US-0168674P.
PR 03-DEC-1999; 99US-0168675P.
PR 07-DEC-1999; 99US-0169278P.
PR 07-DEC-1999; 99US-0169298P.
PR 07-DEC-1999; 99US-0169302P.
PR 08-DEC-1999; 99US-0169691P.
PR 08-DEC-1999; 99US-0169692P.
PR 16-DEC-1999; 99US-0171098P.
PR 16-DEC-1999; 99US-0171107P.
PR 16-DEC-1999; 99US-0171114P.
PR 19-JAN-2000; 2000US-0176866P.
PR 19-JAN-2000; 2000US-0176867P.
PR 19-JAN-2000; 2000US-0176910P.
PR 26-JAN-2000; 2000US-0178166P.
PR 27-JAN-2000; 2000US-0177666P.
PR 27-JAN-2000; 2000US-0178544P.
PR 27-JAN-2000; 2000US-0178545P.
PR 27-JAN-2000; 2000US-0178546P.
PR 27-JAN-2000; 2000US-0178547P.
PR 28-JAN-2000; 2000US-0178754P.
PR 28-JAN-2000; 2000US-0178755P.
PR 01-FEB-2000; 2000US-0179388P.
PR 01-FEB-2000; 2000US-0179395P.
PR 03-FEB-2000; 2000US-0180039P.
PR 03-FEB-2000; 2000US-0180139P.
PR 04-FEB-2000; 2000US-0180206P.
PR 04-FEB-2000; 2000US-0180207P.
PR 07-FEB-2000; 2000US-0180695P.
PR 07-FEB-2000; 2000US-0180696P.
PR 09-FEB-2000; 2000US-0181214P.
PR 09-FEB-2000; 2000US-0181228P.
PR 10-FEB-2000; 2000US-0181476P.
PR 10-FEB-2000; 2000US-0181551P.
PR 15-FEB-2000; 2000US-0182477P.
PR 15-FEB-2000; 2000US-0182478P.
PR 15-FEB-2000; 2000US-0182512P.
PR 15-FEB-2000; 2000US-0182516P.
PR 17-FEB-2000; 2000US-0183165P.
PR 17-FEB-2000; 2000US-0183166P.
XX
PA (CERE-) CERES INC.
XX
PI Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L, Dumas J;
XX
DR WPI; 2000-507395/46.
DR N-PSDB; AAC50846.
XX
PT New sequence determined DNA fragments (SDFs) from different plant
PT species, e.g. corn, rice or Arabidopsis thaliana, useful as promoters,
PT protein coding sequences, untranslated regions, or as 3' termination
PT sequences.
XX
PS Claim 19; SEQ ID NO 66354; 344pp + Sequence Listing; English.
XX
CC The present sequence is a putative protein fragment from Arabidopsis
CC thaliana. Its coding sequence was isolated by carrying out RT-PCR on all
CC of the mRNA obtained from the plant, and creating a cDNA library which
CC could then be sequenced, allowing the putative protein sequence(s) to be
CC obtained. This sequence may be useful for protein identification and for
CC aiding in the elucidation of signal transduction and metabolic pathways.
CC Its coding sequence has a use in the control of gene expression as a
CC promoter, coding sequence, 3'UTR or termination sequence, for controlling
CC the behaviour of a gene within the chromosome, as a tool for use in
CC genetic mapping, including a use in hybridisation assays, for recognition
CC or isolation of similar DNA fragments, or for the identification of a
CC particular organism
XX
SQ Sequence 151 AA;

AAG52218 Length: 151 October 5, 2004 14:15 Type: P Check: 9159 ..

1 MPCSSDHEAW MKCYKENIGS PLKCSGFVKS FQDCARRSRQ QVNPEENSNT

51 LNRVNLGEQI FLSIFNVMTR MMLGAIVEEE ERTILGNELK KLILLFQISK

101 EAQKYSYVPH RNSKTSVVAG YTVLKDLKIF INVWAINRDT KNWEEEGSKR

151 V

!!AA_SEQUENCE 1.0
ID **AAG64458** standard; protein; 542 AA.
XX
AC AAG64458;
XX
DT 22-OCT-2001 (first entry)
XX
DE S. putrefaciens eicosapentanoic acid synthesis enzyme 6.
XX
KW Cyanobacterium; eicosapentanoic acid; EPA; plasmid.
XX
OS Shewanella putrefaciens.
XX
PN JP2001145490-A.
XX
PD 29-MAY-2001.
XX
PF 19-NOV-1999; 99JP-00329169.
XX
PR 19-NOV-1999; 99JP-00329169.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (BIOI-) BIOINDUSTRY KYOKAI SH.
PA (KEIZ-) KEIZAI SANGYOSHO SANGYO GIJUTSU SOGO KEN.
XX
DR WPI; 2001-406151/43.
DR N-PSDB; AAH47839.
XX
PT A plasmid in which eicosapentaenoic acid biosynthesis gene group is
PT cloned and used to transform cyanobacterium so that it produces
PT eicosapentaenoic acid.
XX
PS Claim 2; Page 60-61; 62pp; Japanese.
XX
CC The invention relates to a plasmid prepared by cloning the Shewanella
CC putrefaciens SCRC-2874 (FERM BP-1625) eicosapentanoic acid (EPA)
CC synthesis gene cluster (AAH47833) into a broad host vector. The plasmid
CC is used to transform cyanobacterium and produce EPA. The present sequence
CC is that of a EPA biosynthesis enzyme of the invention
XX
SQ Sequence 542 AA;

```
AAG64458  Length: 542  October 5, 2004 14:15  Type: P  Check: 9499  ..

       1  MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRAC YVVNHADHGF

      51  GIAQTADIVT EQAANSTDLP VSAFTPALGT ESLGDNNFRR VHGVKYAYYA

     101  GAMANGISSE ELVIALGQAG ILCSFGAAGL IPSRVEAAIN RIQAALPNGP

     151  YMFNLIHSPS EPALERGSVE LFLKHKVRTV EASAFLGLTP QIVYYRAAGL

     201  SRDAQGKVVV GNKVIAKVSR TEVAEKFMMP APAKMLQKLV DDGSITAEQM

     251  ELAQLVPMAD DITAEADSGG HTDNRPLVTL LPTILALKEE IQAKYQYDTP

     301  IRVGCGGGVG TPDAALATFN MGAAYIVTGS INQACVEAGA SDHTRKLLAT

     351  TEMADVTMAP AADMFEMGVK LQVVKRGTLF PMRANKLYEI YTRYDSIEAI

     401  PLDEREKLEK QVFRSSLDEI WAGTVAHFNE RDPKQIERAE GNPKRKMALI

     451  FRWYLGLSSR WSNSGEVGRE MDYQIWAGPA LGAFNQWAKG SYLDNYQDRN

     501  AVDLAKHLMY GAAYLNRINS LTAQGVKVPA QLLRWKPNQR MA

!!AA_SEQUENCE 1.0
ID   AAR22511 standard; protein; 488 AA.
XX
AC   AAR22511;
XX
DT   28-JUL-1992  (first entry)
XX
DE   Human Factor Xai.
XX
KW   Mutant; prothrombinase complex; proteolylic; precursor; thrombosis;
KW   inflammation; restenosis; transplantation; haemophilia; antibodies.
XX
OS   Homo sapiens.
XX
PN   WO9204378-A.
XX
PD   19-MAR-1992.
XX
PF   04-SEP-1991;   91WO-US006337.
XX
PR   04-SEP-1990;   90US-00578646.
XX
PA   (CORT-) COR THERAPEUTICS IN.
XX
PI   Wolf D;
XX
DR   WPI; 1992-114303/14.
XX
PT   New analogues of Factor XA peptide - useful for treating haemophilia,
PT   thrombosis, inflammation and transplant complications, for in=vivo
PT   diagnosis.
XX
PS   Claim 3; Fig 1; 59pp; English.
XX
CC   The full length cDNA for factor X (see Leytus, S.P. et al., Proc. Natl.
CC   Acad. Sci. USA (1984) 81: 3699) was subcloned in M13mp19 and subjected to
CC   site specific mutagenesis to replace Ser 185 and Asp 88. In a typical
CC   case mutagenesis was first performed with an oligonucleotide extending
CC   Arg 142 of the light chain by Arg and Lys, then aligning with Ile 53 of
CC   the activation peptide. When expressed in CHO cells the truncated peptide
CC   was cleaved endogenously. Modified Factor Xa was further produced by
CC   acylation e.g. with the p-nitrophenyl ester of p-toluoylic acid. Factor
CC   Xai is used to treat or prevent thrombosis; inflammation; resenosis or
CC   complications of transplantation. It is also used in treatment of adult
CC   respiratory distress syndrome and haemophilia. The modified Factor Xai
CC   has no proteolytic activity and interferes with the ability of endogenous
CC   factor Xa to convert prothrombin to thrombin. Antibodies reactive with
CC   Factor Xai are passive therapeutic agents and used for diagnosis. See
CC   also AAR22513
XX
SQ   Sequence 488 AA;

 AAR22511  Length: 488  October 5, 2004 14:12  Type: P  Check: 2459  ..

       1  MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

      51  GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

     101  CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS

     151  CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW

     201  KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA

     251  LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD RNTEQEEGGE

     301  AVHEVEVVIK HNRFTKETYD FDIAVLRLKT PITFRMNVAP ACLPERDWAE

     351  STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDRNSC KLSSSFIITQ

     401  NMFCAGYDTK QEDACQGDSG GPHVTRFKDT YFVTGIVSWG EGCARKGKYG

     451  IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
ID   AAR22512 standard; protein; 488 AA.
XX
AC   AAR22512;
XX
DT   28-JUL-1992  (first entry)
XX
DE   Mutated precursor of human Factor X analogue.
XX
KW   Mutant; prothrombinase complex; proteolylic; precursor; thrombosis;
KW   inflammation; restenosis; transplantation; haemophilia; antibodies.
XX
OS   Homo sapiens.
XX
PN   WO9204378-A.
XX
PD   19-MAR-1992.
XX
PF   04-SEP-1991;   91WO-US006337.
XX
PR   04-SEP-1990;   90US-00578646.
XX
PA   (CORT-) COR THERAPEUTICS IN.
XX
PI   Wolf D;
XX
DR   WPI; 1992-114303/14.
XX
PT   New analogues of Factor XA peptide - useful for treating haemophilia,
PT   thrombosis, inflammation and transplant complications, for in=vivo
PT   diagnosis.
XX
PS   Claim 3; Fig 1; 59pp; English.
XX
CC   The full length cDNA of human factor X was obtd. from Dr. W.R. Church,
CC   University of Vermont. This human Factor X cDNA was cloned into the EcoRI
CC   site of vector pBSII (Stratagene) to obtain pBSX. The HindIII-XbaI
CC   fragment of pBSX comprising the entire Factor X coding region was
CC   subcloned into the HindIII-XbaI site of vector M13mp19 (Mp19X).
CC   Oligonucleotide site-directed mutagenesis was then performed using
CC   oligomers to convert serine 185 on the Factor X heavy chain to alanine,
CC   and to convert aspartic acid 88 on the Factor X heavy chain to
CC   asparagine. Modified Factor Xa was further produced by acylation e.g.
CC   with the p-nitrophenyl ester of p-toluoylic acid. Modified Factor X is
CC   used to treat or prevent thrombosis; inflammation; restenosis or
CC   complications of transplantation. It is also used in treatment of adult
CC   respiratory distress syndrome and haemophilia. The modified Factor X has
CC   no proteolytic activity and interferes with the ability of endogenous
```

```
CC   factor Xa to convert prothrombin to thrombin. Antibodies reactive with
CC   modified Factor X are passive therapeutic agents and used for diagnosis.
CC   See also AAR22513
XX
SQ   Sequence 488 AA;

 AAR22512  Length: 488  October 5, 2004 14:12  Type: P  Check: 2469  ..

       1  MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

      51  GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

     101  CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS

     151  CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW

     201  KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA

     251  LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD RNTEQEEGGE

     301  AVHEVEVVIK HNRFTKETYD FNIAVLRLKT PITFRMNVAP ACLPERDWAE

     351  STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDRNSC KLSSSFIITQ

     401  NMFCAGYDTK QEDACQGDAG GPHVTRFKDT YFVTGIVSWG EGCARKGKYG

     451  IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
ID   AAR30729 standard; protein; 870 AA.
XX
AC   AAR30729;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   20-MAY-1993  (first entry)
XX
DE   p100 protein from human herpes virus type 6.
XX
KW   antibodys; monoclonal antibody; ELISA assays; CMV; cytomegalovirus.
XX
OS   Human herpesvirus 6.
XX
PN   EP524421-A1.
XX
PD   27-JAN-1993.
XX
PF   15-JUN-1992;   92EP-00110047.
XX
PR   08-JUL-1991;   91EP-00111338.
XX
PA   (BEHW ) BEHRINGWERKE AG.
XX
PI   Neipel F,  Fleckenstein B;
XX
DR   WPI; 1993-028531/04.
DR   P-PSDB; AAR30729.
XX
PT   Human herpes virus type 6 protein p100 DNA sequence - useful in
PT   prophylaxis, treatment and differential diagnosis of human herpes virus-6
PT   infections.
XX
PS   Claim 1; Page 12; 25pp; English.
XX
CC   This sequence is the p100 protein from human herpes virus type6. The
CC   protein and antibodies to it can be used for treatment or prevention of
CC   HHV-6 infections. The DNA , protein and Ab are also useful in eg. ELISA
CC   assays esp. for differentiating between HHV-6 and cytomegalovirus
CC   infections. These assays are more sensitive and specific than
CC   immunofluorescence methods currently used. (Updated on 25-MAR-2003 to
CC   correct PN field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 870 AA;

 AAR30729  Length: 870  October 5, 2004 14:12  Type: P  Check: 500  ..

       1  MDLQRHPIPF AWLDRDKVER LTDFLSNLER LDNVDLREHP HVTNSCVVRE

      51  GDDVDDLKTL YNLLVLWLMY HYVLSKRKPD YNAIWQDITK LQSVVNEYLN

     101  SKGLNKGIFE NMFTNKEKFE SQFSDINRAL LRLGNFIKWG SNVAIDTPYV

     151  NLTAEDSSEI ENNLQDAEKN MLWYTVYNIN DPWDENGYLI TSINKLIYLG

     201  KLFLALTQSW SKLEKVAMSQ IVITQNHLSG HLRRHDNFNI VYSHRVLQTP

     251  LTGQRVESFL KIITSDYDII KSSLESHSAS KAFSMSEIGP NSLMDFVPLR

     301  GDIHSNLTLP SMSIDTKKSS LDPARLKKSN SRSLDSFLRM QRQPKFLELD

     351  SVDNAGEKIL LKEATLGGEN VKATTPASSV SLMSGVESPS SFTSTNLDLP

     401  LSSFTSTNLD LRDKSHGNYK IGPSGILDFN VKFPPNAQLN TNGVDLLQDK

     451  TSIGSPSSGI TDVVNGFANL NLHQNKSNVS PPWSRNTAAN ADFLDPVHRF

     501  VPEQTGTPFV LNNSDVAGSE AKHTTYSTET GVSPRNVFLI KDLRGKDGFR

     551  KQKQSDIPKS LTKERNDKAI MHSREVTGDS GDATETVGAR NSPALRKIKQ

     601  ANDFFAGLNK KNDRDVLRGG KGNSKDLHSG GNAKKKEMSG KFNDDKEMTR

     651  NGQEPSRSLM GDARNAGDEQ YIQAGLGQRV NNLLSQFTNL ISLGEKGIED

     701  ILQNQRGTEL KLATENKSGR ESEEANVEKI LEVSNPQDMF KNFRLQNDLD

     751  SVQSPFRLPD ADLSRELDSA SFKDALDLKL PGNGEREIDL ALEKVKVGET

     801  ETSDLKVGQD ESFVPAQLMK VETPEEKDDI IEQMVLRIRQ DGETDENTVS

     851  GPGVAESLDI EAKGESAIAS

!!AA_SEQUENCE 1.0
ID   AAR35762 standard; protein; 448 AA.
XX
AC   AAR35762;
XX
DT   25-MAR-2003  (revised)
DT   24-SEP-1993  (first entry)
XX
DE   Factor X (X).
XX
KW   PC; protein C; IX; Factor IX; X; Factor X; PT; prothrombin; VII;
KW   Factor VII; CT; chymotrypsinogen; SP; serine protease; binding; exosite;
KW   catalytic activity.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .139
FT                   /note= "Factor X light chain"
FT   Region          140. .142
FT                   /note= "Factor X activation"
FT   Region          143. .448
FT                   /note= "Factor X heavy chain"
FT   Peptide         285. .306
FT                   /note= "claim 7, page 138 describes an antibody that
FT                   reacts with Factor X; fragments 330-344, 404. .418 and 415
FT                   -429 but not with fragment 285-306"
FT   Peptide         330. .350
FT                   /note= "exosite 2"
FT   Peptide         330. .344
FT                   /note= "pref. PC polypeptide; claim 2, page 136"
FT   Peptide         404. .418
FT                   /note= "pref. PC polypeptide; claim 2, page 136"
FT   Peptide         409. .423
```

FT /note= "exosite 1"
FT Peptide 415. .429
FT /note= "pref. PC polypeptide; claim 2, page 136"
XX
PN WO9309804-A1.
XX
PD 27-MAY-1993.
XX
PF 18-NOV-1992; 92WO-US010242.
XX
PR 18-NOV-1991; 91US-00793989.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Griffin JH, Mesters RM;
XX
DR WPI; 1993-182244/22.
XX
PT Serine protease derived-polypeptide(s) and anti-peptide antibodies - for
PT inhibiting coagulation and assaying for the presence of serine protease
PT in fluid samples.
XX
PS Disclosure; Page 128-130; 149pp; English.
XX
CC The PC polypeptides indicated in the Features Table inhibit coagulation
CC (they prevent binding of serine protease to natural substrates), esp.
CC when admin. to give an intravascular blood concn. of 0.1-100 (pref. 0.5-
CC 10) microM. NB: Sequences corresp. to SEQ ID NO 6, 7, 8 and 9 are
CC described in the specification but have not yet been added to the
CC SEQUENCE LISTING. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 448 AA;

AAR35762 Length: 448 October 5, 2004 14:13 Type: P Check: 40 ..

1 ANSFLEEMKK GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC

51 ETSPCQNQGK CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC

101 HEEQNSVVCS CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS

151 SGEAPDSITW KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE

201 CKDGECPWQA LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD

251 RNTEQEEGGE AVHEVEVVIK HNRFTKETYD FDIAVLRLKT PITFRMNVAP

301 ACLPERDWAE STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDRNSC

351 KLSSSFIITQ NMFCAGYDTK QEDACQGDSG GPHVTRFKDT YFVTGIVSWG

401 EGCARKGKYG IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
ID **AAR42456** standard; protein; 543 AA.
XX
AC AAR42456;
XX
DT 25-MAR-2003 (revised)
DT 27-MAY-1994 (first entry)
XX
DE Enzyme involved in eicosapentaenoic acid (EPA) synthesis.
XX
KW EPA; eicosapentaenoic acid synthetase; drug; anticoagulant; hypolipemic;
KW hypoglycemic; antihypertensive; anticancer; pesticide; foodstuff;
KW additive.
XX
OS Shewanella putrefaciens.
XX
PN WO9323545-A1.
XX
PD 25-NOV-1993.
XX
PF 14-MAY-1993; 93WO-JP000641.
XX
PR 15-MAY-1992; 92JP-00147945.
XX
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
PI Yazawa K, Yamada A, Kato S, Kondo K;
XX
DR WPI; 1993-386577/48.
DR N-PSDB; AAQ51128.
XX
PT Gene coding for eicosa-penta:enoic acid synthetase - is isolated from
PT Pseudomonas, Alteromonas or Shewanella and used for recombinant prodn. of
PT eicosa-penta:enoic acid.
XX
PS Claim 10; Page 91-94; 106pp; Japanese.
XX
CC EPA is useful as a drug, having anticoagulant, hypolipemic, hypoglycemic,
CC antihypertensive and anticancer activity. It is also a pesticide and is
CC useful as a nutritional foodstuff and animal feed additive. (Updated on
CC 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 543 AA;

AAR42456 Length: 543 October 5, 2004 14:12 Type: P Check: 3609 ..

1 MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRAC YVVNHADHGF

51 GIAQTADIVT EQAANSTDLP VSAFTPALGT ESLGDNNFRR VHGVKYAYYA

101 GAMANGISSE ELVIALGQAG ILCGSFGAAG LIPSRVEAAI NRIQAALPNG

151 PYMFNLIHSP SEPALERGSV ELFLKHKVRT VEASAFLGLT PQIVYYRAAG

201 LSRDAQGKVV VGNKVIAKVS RTEVAEKFMM PAPAKMLQKL VDDGSITAEQ

251 MELAQLVPMA DDITAEADSG GHTDNRPLVT LLPTILALKE EIQAKYQYDT

301 PIRVGCGGGV GTPDAALATF NMGAAYIVTG SINQACVEAG ASDHTRKLLA

351 TTEMADVTMA PAADMFEMGV KLQVVKRGTL FPMRANKLYE IYTRYDSIEA

401 IPLDEREKLE KQVFRSSLDE IWAGTVAHFN ERDPKQIERA EGNPKRKMAL

451 IFRWYLGLSS RWSNSGEVGR EMDYQIWAGP ALGAFNQWAK GSYLDNYQDR

501 NAVDLAKHLM YGAAYLNRIN SLTAQGVKVP AQLLRWKPNQ RMA

!!AA_SEQUENCE 1.0
ID **AAR99465** standard; protein; 543 AA.
XX
AC AAR99465;
XX
DT 30-JAN-1997 (first entry)
XX
DE Biosynthetic enzyme of icosapentaenoic acid synthase.
XX
KW Icosapentaenoic acid synthase; EPA; drugs; agrochemicals; foodstuffs;
KW animal feed; lipid balance correction; antihypertensive;
KW antiinflammatory; anticancer agent.
XX
OS Shewanella putrefaciens.
XX
PN WO9621735-A1.
XX
PD 18-JUL-1996.
XX
PF 12-JAN-1996; 96WO-JP000030.
XX
PR 13-JAN-1995; 95JP-00004299.
XX
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX

PI Yazawa K, Yamada A, Kondo K;
XX
DR WPI; 1996-342288/34.
DR N-PSDB; AAT34137.
XX
PT Production of icosapentaenoic acid using transformed E. coli - uses DNA
PT coding for icosapentaenoic acid synthase derived from Shewanella strain.
XX
PS Claim 7; Page 128-131; 145pp; English.
XX
CC The DNA sequence (AAT34137) which encodes the biosynthetic enzymes of
CC icosapentaenoic acid (EPA) can be used to transform Escherichia coli. The
CC DNA sequence allows efficient microbial production of EPA, which is a raw
CC material for drugs, agrochemicals, foods and animal feedstuffs. EPA is
CC also useful for lipid balance correction and as an antihypertensive,
CC antiinflammatory and anticancer agent
XX
SQ Sequence 543 AA;

AAR99465 Length: 543 October 5, 2004 14:13 Type: P Check: 3609 ..

1 MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRAC YVVNHADHGF

51 GIAQTADIVT EQAANSTDLP VSAFTPALGT ESLGDNNFRR VHGVKYAYYA

101 GAMANGISSE ELVIALGQAG ILCGSFGAAG LIPSRVEAAI NRIQAALPNG

151 PYMFNLIHSP SEPALERGSV ELFLKHKVRT VEASAFLGLT PQIVYYRAAG

201 LSRDAQGKVV VGNKVIAKVS RTEVAEKFMM PAPAKMLQKL VDDGSITAEQ

251 MELAQLVPMA DDITAEADSG GHTDNRPLVT LLPTILALKE EIQAKYQYDT

301 PIRVGCGGGV GTPDAALATF NMGAAYIVTG SINQACVEAG ASDHTRKLLA

351 TTEMADVTMA PAADMFEMGV KLQVVKRGTL FPMRANKLYE IYTRYDSIEA

401 IPLDEREKLE KQVFRSSLDE IWAGTVAHFN ERDPKQIERA EGNPKRKMAL

451 IFRWYLGLSS RWSNSGEVGR EMDYQIWAGP ALGAFNQWAK GSYLDNYQDR

501 NAVDLAKHLM YGAAYLNRIN SLTAQGVKVP AQLLRWKPNQ RMA

!!AA_SEQUENCE 1.0
ID **AAU39422** standard; protein; 491 AA.
XX
AC AAU39422;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #318.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59507.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID NO 617; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 491 AA;

AAU39422 Length: 491 October 5, 2004 14:16 Type: P Check: 2758 ..

1 YVEYGPVAQR IEQQPSNLSR VGSIPTRAAT VSSRVSPGED LAHRKVSRAA

51 GSLSTMSYVA PTPIARRPWW QRALRNPWFW VLLALVLISA ACMVHTYVIV

101 HADTEVEQQG QKGVIPGITN QSLKLAFHYA WPTAAVWSAI FIFLDRFRTR

151 HLHVWFLCFC WGACVATWIS MHVNTWMAGM LSVTGGVDPA SGAGPAVYSA

201 PFVEESCKAL VLFALAIGMG RRMTSVVQTV SMAGLSAIGF AFVENIMYYA

251 RADNYARVTA SAGDPKQAVM ELVLLRGVYA SFGHPLFTSM TGIGLALGLR

301 SRSRLVRIFA PTTGFVMAVV GHMLFNGFSS VLPMAILKKL WFVALGIVAS

351 VVVFLVVRTV REGRMIRYRL EDYVKGGWLP NSDADTLSAL RRRQWAALVA

401 LSQGPRRWWH TLEFLRVGTD LAYLRDAIVR GLDDDPGPRQ IELINRMNVL

451 RPAAITVARG AKLSKPRLFA FLKRRRNQPV NNELQWAPPQ A

!!AA_SEQUENCE 1.0
ID **AAW05820** standard; protein; 306 AA.
XX
AC AAW05820;
XX
DT 17-DEC-1996 (first entry)
XX
DE Factor X heavy chain.
XX
KW Factor X; light chain; human; heavy chain; Factor Xa; prothrombinase;
KW activating enzyme; blood factor; immunoaffinity chromatography; therapy;
KW antigen; inhibitor; activated resin; coagulation disorder; factor II;
KW vasculature function disorder; factor V; factor IX; factor XI; protein C;
KW factor XII; factor VII; protein S; fibrinogen.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

```
FT   Peptide          1. .52
FT                    /note= "activator peptide"
FT   Disulfide-bond   59. .64
FT   Disulfide-bond   79. .95
FT   Disulfide-bond   160
FT                    /note= "disulphide bond to residue 132 of Factor X light
FT                    chain (see AAR95596)"
FT   Disulfide-bond   208. .222
FT   Disulfide-bond   233. .261
XX
PN   WO9613274-A1.
XX
PD   09-MAY-1996.
XX
PF   27-OCT-1995;   95WO-US013940.
XX
PR   28-OCT-1994;   94US-00330978.
XX
PA   (CORT-) COR THERAPEUTICS INC.
XX
PI   King RS;
XX
DR   WPI; 1996-239270/24.
XX
PT   Prepn. of an inhibited form of an activated blood factor, e.g. factor X -
PT   by treating partially purified blood factor prepns. with an activating
PT   factor and an inhibiting factor.
XX
PS   Disclosure; Fig 1; 45pp; English.
XX
CC   This sequence represents the heavy chain of human Factor X (the light
CC   chain is represented by AAR95596). Factor X must be activated to Factor
CC   Xa before the protease is incorporated into the prothrombinase complex.
CC   In Factor Xa the light chain sequence is identical to the Factor X light
CC   chain, and the heavy chain is a truncated version of the Factor X heavy
CC   chain. An inhibited form of activated Factor X is prepared by the method
CC   of the invention. In this method, a partially purified preparation
CC   containing the blood factor is treated to convert the factor into an
CC   activated form (using an immobilised activating enzyme). The activated
CC   form is then converted into an inhibited form in a single step, and the
CC   inhibited factor is recovered. The inhibited blood factor is recovered by
CC   immunoaffinity chromatography using an antigen specific monoclonal
CC   antibody coupled to an activated resin (such as agarose), or an anion
CC   exchange column with an anion-exchange group linked to a naturally
CC   derived polysaccharide or a synthetically derived polymeric matrix. The
CC   activated resin used preferably uses activation chemistry selected from
CC   tresyl, azactone, aldehyde, hydrazide, N-hydroxy succinimide or triazine.
CC   This method produces a highly purified preparation of an inhibited form
CC   (either permanently or transiently inhibited) of an activated blood
CC   factor in high yield. The factors produced can be used in the treatment
CC   of coagulation disorders, or disorders of vasculature function
XX
SQ   Sequence 306 AA;

 AAW05820  Length: 306  October 5, 2004 14:13  Type: P  Check: 8643  ..

       1  SVAQATSSSG EAPDSITWKP YDAADLDPTE NPFDLLDFNQ TQPERGDNNL

      51  TRIVGGQECK DGECPWQALL INEENEGFCG GTILSEFYIL TAAHCLYQAK

     101  RFKVRVGDRN TEQEEGGEAV HEVEVVIKHN RFTKETYDFD IAVLRLKTPI

     151  TFRMNVAPAC LPERDWAEST LMTQKTGIVS GFGRTHEKGR QSTRLKMLEV

     201  PYVDRNSCKL SSSFIITQNM FCAGYDTKQE DACQGDSGGP HVTRFKDTYF

     251  VTGIVSWGEG CARKGKYGIY TKVTAFLKWI DRSMKTRGLP KAKSHAPEVI

     301  TSSPLK

!!AA_SEQUENCE 1.0
ID   AAW37053 standard; protein; 543 AA.
XX
AC   AAW37053;
XX
DT   03-JUL-1998  (first entry)
XX
DE   S. putrefaciens EPO biosynthesis gene cluster ORF9 product.
XX
KW   SCRC-2874; FERM BP-1625; eicosapentaenoic acid; EPA;
KW   biosynthesis gene cluster; synthetase.
XX
OS   Shewanella putrefaciens.
XX
PN   WO9801565-A1.
XX
PD   15-JAN-1998.
XX
PF   09-JUL-1997;   97WO-JP002371.
XX
PR   10-JUL-1996;   96JP-00180845.
XX
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI   Yazawa K,  Yamada A,  Kondo K,  Kato S;
XX
DR   WPI; 1998-101060/09.
DR   N-PSDB; AAV00503.
XX
PT   Eicosapentaenoic acid produced by culture of transformed Escherichia coli
PT   - containing an eicosapentaenoic acid synthetase gene derived from the
PT   marine microorganism Shewanella.
XX
PS   Example 1; Page 93-97; 110pp; Japanese.
XX
CC   The present sequence is encoded by the Shewanella putrefaciens SCRC-2874
CC   (FERM BP-1625) eicosapentaenoic acid (EPA) biosynthesis gene cluster. A
CC   novel EPA (useful in drugs, pesticides, foods and feedstuffs) is encoded
CC   by synthetase enzyme gene sequences comprising parts of the full sequence
CC   of the synthetase gene from the marine microorganism S. putrefaciens SCRC
CC   -2874 (FERM BP-1625), in which at least 1 of the 9 open reading frames
CC   (ORF) (numbered 2-10) in the gene have been deleted. In particular the
CC   gene sequences comprising the following parts of the full gene: (1) bases
CC   8081-9441, 12314-13084 and 13889-32520; (2) bases 8081-9441, 12314-13084,
CC   13889-32520 and 34627-35559; (3) bases 8081-9441, 12314-13084 and 13889-
CC   35559; (4) bases 8081-9441, 9681-13084 and 13889-32520; (5) bases 8081-
CC   9441, 9681-13084, 13889-32520 and 34627-35564; and (6) bases 8081-9441,
CC   9681-13084 and 13889-35564, are claimed
XX
SQ   Sequence 543 AA;

 AAW37053  Length: 543  October 5, 2004 14:14  Type: P  Check: 3609  ..

       1  MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRAC YVVNHADHGF

      51  GIAQTADIVT EQAANSTDLP VSAFTPALGT ESLGDNNFRR VHGVKYAYYA

     101  GAMANGISSE ELVIALGQAG ILCGSFGAAG LIPSRVEAAI NRIQAALPNG

     151  PYMFNLIHSP SEPALERGSV ELFLKHKVRT VEASAFLGLT PQIVYYRAAG

     201  LSRDAQGKVV VGNKVIAKVS RTEVAEKFMM PAPAKMLQKL VDDGSITAEQ

     251  MELAQLVPMA DDITAEADSG GHTDNRPLVT LLPTILALKE EIQAKYQYDT

     301  PIRVGCGGGV GTPDAALATF NMGAAYIVTG SINQACVEAG ASDHTRKLLA

     351  TTEMADVTMA PAADMFEMGV KLQVVKRGTL FPMRANKLYE IYTRYDSIEA

     401  IPLDEREKLE KQVFRSSLDE IWAGTVAHFN ERDPKQIERA EGNPKRKMAL

     451  IFRWYLGLSS RWSNSGEVGR EMDYQIWAGP ALGAFNQWAK GSYLDNYQDR

     501  NAVDLAKHLM YGAAYLNRIN SLTAQGVKVP AQLLRWKPNQ RMA

!!AA_SEQUENCE 1.0
```

ID **AAW40283** standard; protein; 467 AA.
XX
AC AAW40283;
XX
DT 16-JUN-1998 (first entry)
XX
DE Human Factor X protease.
XX
KW Factor X; factor IX; serine protease activity; catalytic domain; ZAD;
KW zymogen-activating domain; epidermal growth factor-like domain; EGF1;
KW EGF2; regulator; coagulation; fibrinolysis; homeostasis; X-ray structure;
KW detection; drug modelling; restriction protease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 108. .153
FT /label= EGF2 domain
FT Domain 154. .165
FT /label= EGF2 domain
FT Domain 166. .216
FT /label= Activating domain
FT Domain 217. .454
FT /label= catalytic domain
XX
PN WO9747737-A1.
XX
PD 18-DEC-1997.
XX
PF 11-JUN-1997; 97WO-EP003027.
XX
PR 11-JUN-1996; 96EP-00109288.
PR 22-JUN-1996; 96EP-00110109.
PR 06-JUL-1996; 96EP-00110959.
XX
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
PI Kopetzki E, Hopfner K;
XX
DR WPI; 1998-052304/05.
DR N-PSDB; AAV10462.
XX
PT Non-glycosylated, truncated forms of factor IX family protein with serine
PT protease activity - used to screen for specific modulators and to assay
PT factor IXa.
XX
PS Disclosure; Fig 3; 49pp; German.
XX
CC This sequence represents a human factor X protease. This protein is used
CC in the construction of a novel non-glycosylated protein and truncated and
CC zymogen forms of this protein, which have serine protease activity. The
CC protein is composed of various domains from a factor IX family protein,
CC namely a catalytic domain (CD) N-terminally bound to a zymogen-activating
CC domain (ZAD), N-terminally bound to an EGF1 and/or EGF2 domain (EGF =
CC epidermal growth factor-like domain). Such proteins are used to identify
CC activators/inhibitors of factor IX family proteins (potentially useful as
CC regulators of coagulation, fibrinolysis and homeostasis). The protein in
CC zymogen form is also useful in assays for detecting factor IXa activity
CC in aqueous solution (specifically in body fluids). The protein can be
CC used to produce co-crystals with protease variants or inhibitors for X-
CC ray structural analysis and drug modelling and as restriction proteases
CC in biotechnology. These truncated proteins have the same specificity as
CC factor IX family proteases and can be produced in prokaryotes in a form
CC that allows production of active enzyme by conversion to native form and
CC enzymatic cleavage
XX
SQ Sequence 467 AA;

AAW40283 Length: 467 October 5, 2004 14:14 Type: P Check: 1666 ..

1 LLGESLFIRR EQANNILARV TRANSFLEEM KKGHLERECM EETCSYEEAR

51 EVFEDSDKTN EFWNKYKDGD QCETSPCQNQ AKCKDGLGEY TCTCLEGFEG

101 KNCELFTRKL CSLDNGDCDQ FCHEEQNSVV CSCARGYTLA DNGKACIPTG

151 PYPCGKQTLE RRKRSVAQAT SSSGEAPDSI TWKPYDAADL DPTENPFDLL

201 DFNQTQPERG DNNLTRIVGG QECKDGECPW QALLINEENE GFCGGTILSE

251 FYILTAAHCL YQAKRFEGDR NTEQEEGGEA VHEVEVVIKH NRFTKETYDF

301 DIAVLRLKTP ITFRMNVAPA CLPERDWAES TLMTQKTGIV SGFGRTHEKG

351 RQSTRLKMLE VPYVDRNSCK LSSSFIITQN MFCAGYDTKQ EDACQGDSGG

401 PHVTRFKDTY FVTGIVSWGE GCARKGNYGI YTKVTAFLKW IDRSMKTRGL

451 PKAKSHAPEV ITSSPLK

!!AA_SEQUENCE 1.0
ID **AAW66092** standard; peptide; 448 AA.
XX
AC AAW66092;
XX
DT 16-NOV-1998 (first entry)
XX
DE Human factor X variant.
XX
KW factor X variant; factor V; fVa; diagnostic assay; heparin; thrombin;
KW blood coagulation.
XX
OS Homo sapiens.
XX
PN WO9839456-A1.
XX
PD 11-SEP-1998.
XX
PF 05-MAR-1998; 98WO-US003939.
XX
PR 07-MAR-1997; 97US-0040047P.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Miletich JP;
XX
DR WPI; 1998-495855/42.
XX
PT New human blood coagulation Factor X variant - with asparagine residue at
PT position 347, has reduced affinity for activated Factor V and is useful
PT for diagnostic assays.
XX
PS Claim 1; Page; 42pp; English.
XX
CC The invention relates to a human Factor X variant (nfX) where asparagine
CC replaces arginine at position 347. Also claimed is a method for
CC substantially reducing the affinity of human Factor X for activated
CC Factor V (fVa) without substantially reducing the catalytic impact of fVa
CC binding, by replacing arginine with asparagine at position 347. The new
CC fX variant is especially useful for in vitro assays and diagnostic
CC applications. Specifically, these include (1) quantifying the importance
CC of the interaction between the serine protease domain of activated Factor
CC X (fXa) and fVa by comparison of wild type fXa and nfX; (2) quantifying
CC the impact of specific inhibitors of fVa-fXa interaction by comparison of
CC wild type fXa and nfX, which is useful for treatment with inhibitors like
CC heparin and tissue factor pathway inhibitor (TFPT); and (3) reactions
CC where total specificity of thrombin activation to membrane surfaces is
CC required, as nfX has no significant interaction with fVa in the absence
CC of a sufficiently charged phospholipid surface. Substitution of wild-type
CC arginine by asparagine at position 347 of factor X selectively attenuates
CC the interaction between fXa and fVa without affecting its catalytic
CC (thrombogenic) activity (except in the presence of sub-saturating heparin
CC where the rate of inhibition by antithrombin III is 15% of normal). The
CC present sequence represents the specifically claimed human Factor X
CC variant having the arginine residue at position 347 replaced with
CC asparagine. NB: This sequence does not appear as such in the present

```
CC   patent specification but was created using the native factor X sequence
CC   as shown in PIR Accession Number 538554
XX
SQ   Sequence 448 AA;

  AAW66092  Length: 448  October 5, 2004 14:14  Type: P  Check: 20  ..

       1  ANSFLEEMKK GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC

      51  ETSPCQNQGK CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC

     101  HEEQNSVVCS CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS

     151  SGEAPDSITW KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE

     201  CKDGECPWQA LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD

     251  RNTEQEEGGE AVHEVEVVIK HNRFTKETYD FDIAVLRLKT PITFRMNVAP

     301  ACLPERDWAE STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDNNSC

     351  KLSSSFIITQ NMFCAGYDTK QEDACQGDSG GPHVTRFKDT YFVTGIVSWG

     401  EGCARKGKYG IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
ID   AAW76216 standard; protein; 488 AA.
XX
AC   AAW76216;
XX
DT   27-NOV-1998  (first entry)
XX
DE   Human Factor X protein.
XX
KW   Factor X; analogue; activation cleavage site; protease; bleeding; human;
KW   Factor IX; Factor VII; Factor VIII; haemophilia; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .40
FT                   /label= signal
FT   Protein         41. .488
FT                   /label= Factor_X
XX
PN   WO9838317-A1.
XX
PD   03-SEP-1998.
XX
PF   27-FEB-1998;   98WO-AT000045.
XX
PR   27-FEB-1997;   97AT-00000335.
XX
PA   (IMMO ) IMMUNO AG.
XX
PI   Himmelspach M,  Schlokat U,  Dorner F,  Fisch A,  Eibl J;
XX
DR   WPI; 1998-481211/41.
DR   N-PSDB; AAV56776.
XX
PT   New factor X analogues with processing site for protease not active on
PT   natural protein - and related DNA, is very stable and can be activated in
PT   vitro or in vivo without using animal protease(s), particularly for
PT   treating disorders of blood coagulation.
XX
PS   Claim 3; Fig 1; 86pp; German.
XX
CC   This sequence represents the human Factor X protein which is used in a
CC   method resulting in the production of novel human Factor X (F10)
CC   analogues. Such analogues have in the region of the natural F10a
CC   activation cleavage site, a modification that creates a processing site
CC   for a protease that does not naturally cleave F10 in this region. The
CC   proteins are used to generate, in vivo or in vitro, F10a analogues that
CC   can be used to control bleeding and for treating defects of factors IX,
CC   VII or VIII, e.g.in haemophiliacs who have developed antibodies to
CC   factors VIII and/or IX. The encoding nucleic acid can be used in gene
CC   therapy of the same conditions. The analogues have high stability and can
CC   be activated without use of animal enzymes such as trypsin. Only
CC   activation is affected, their activity is the same as the natural factor.
CC   The analogues can be isolated as a pure single-chain pro-protein (not
CC   usually possible because of rapid processing of the native precursor) and
CC   this converted to two-chain form by subsequent activation. Activated
CC   analogues have good stability and structural integrity and are
CC   practically free of inactive intermediates and autoproteolytic
CC   decomposition products
XX
SQ   Sequence 488 AA;

 AAW76216  Length: 488  October 5, 2004 14:13  Type: P  Check: 2975  ..

       1  MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

      51  GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

     101  CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS

     151  CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW

     201  KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA

     251  LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD RNTEQEEGGE

     301  AVHEVEVVIK HNRFTKETYD FDIAVLRLKT PITFRMNVAP ACLPERDWAE

     351  STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDRNSC KLSSSFIITQ

     401  NMFCAGYDTK QEDACQGDSG GPHVTRFKDT YFVTGIVSWG ESCARKGKYG

     451  IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
ID   AAW76218 standard; protein; 488 AA.
XX
AC   AAW76218;
XX
DT   27-NOV-1998  (first entry)
XX
DE   Human Factor X protein.
XX
KW   Factor X; analogue; activation cleavage site; protease; bleeding; human;
KW   Factor IX; Factor VII; Factor VIII; haemophilia; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .40
FT                   /label= signal
FT   Protein         41. .488
FT                   /label= Factor_X
XX
PN   WO9838318-A1.
XX
PD   03-SEP-1998.
XX
PF   27-FEB-1998;   98WO-AT000046.
XX
PR   27-FEB-1997;   97AT-00000336.
XX
PA   (IMMO ) IMMUNO AG.
XX
PI   Himmelspach M,  Pfleiderer M,  Falkner F,  Eibl J,  Dorner F;
PI   Schlokat U;
XX
DR   WPI; 1998-481212/41.
DR   N-PSDB; AAV56821.
XX
```

PT New factor 10 deletion mutants lacking the natural protease processing
PT site - having a non-natural site inserted, for in vitro activation to
PT products used to treat blood coagulation disorders.
XX
PS Claim 3; Fig 1; 82pp; German.
XX
CC This sequence represents the human Factor X protein which is used in a
CC method resulting in the production of novel human Factor X (F10)
CC analogues. Such analogues have in the region of the natural F10a
CC activation cleavage site, a modification that creates a processing site
CC for a protease that does not naturally cleave F10 in this region. The
CC proteins are used to generate, in vivo or in vitro, F10a analogues that
CC can be used to control bleeding and for treating defects of factors IX,
CC VII or VIII, e.g.in haemophiliacs who have developed antibodies to
CC factors VIII and/or IX. The encoding nucleic acid can be used in gene
CC therapy of the same conditions. The analogues have high stability and can
CC be activated without use of animal enzymes such as trypsin. Only
CC activation is affected, their activity is the same as the natural factor.
CC The analogues can be isolated as a pure single-chain pro-protein (not
CC usually possible because of rapid processing of the native precursor) and
CC this converted to two-chain form by subsequent activation. Activated
CC analogues have good stability and structural integrity and are
CC practically free of inactive intermediates and autoproteolytic
CC decomposition products
XX
SQ Sequence 488 AA;

AAW76218 Length: 488 October 5, 2004 14:14 Type: P Check: 2975 ..

1 MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

51 GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

101 CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS

151 CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW

201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA

251 LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD RNTEQEEGGE

301 AVHEVEVVIK HNRFTKETYD FDIAVLRLKT PITFRMNVAP ACLPERDWAE

351 STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDRNSC KLSSSFIITQ

401 NMFCAGYDTK QEDACQGDSG GPHVTRFKDT YFVTGIVSWG ESCARKGKYG

451 IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
ID **AAW89403** standard; protein; 543 AA.
XX
AC AAW89403;
XX
DT 19-OCT-1999 (first entry)
XX
DE S. putrefaciens PKS-like cluster ORF 9 protein.
XX
KW Polyketide-like synthesis; PKS; PKS-like gene; PUFA; DHA; transgenic;
KW poly-unsaturated fatty acid; eicosapentenoic acid; docosahexanoic acid;
KW EPA; oil; dietary supplement; infant feeeding formulation; malnutrition;
KW intravenous feeding formulation; cooking oil; fat; anti-inflammatory;
KW cholesterol; open reading frame; ORF.
XX
OS Shewanella putrefaciens.
XX
PN WO9855625-A1.
XX
PD 10-DEC-1998.
XX
PF 04-JUN-1998; 98WO-US011639.
XX
PR 04-JUN-1997; 97US-0048650P.
XX
PA (CALJ) CALGENE LLC.
XX
PI Facciotti D, Metz JG, Lassner M;
XX
DR WPI; 1999-070271/06.
DR N-PSDB; AAZ00331.
XX
PT New nucleic acid encoding polyketide-like synthesis enzymes of Vibrio
PT marinus - and transformed plants and microbes that produce
PT polyunsaturated fatty acids, useful as pharmaceuticals and food
PT supplements.
XX
PS Claim 17; Fig 4; 153pp; English.
XX
CC The invention provides polyketide-like synthesis (PKS)-like genes that
CC are used for the production of long chain poly-unsaturated fatty acid
CC (PUFA) productions. Genes responsible for eicosapentenoic acid (EPA)
CC production in Shewanella putrefaciens and novel genes associated with the
CC production of docosahexanoic acid (DHA) in Vibrio marinus are used to
CC generate transgenic plants that can express transgenes encoding PKS-like
CC genes associated with PUFA production. The PKS-like genes are used to
CC transform plants and microbial cells to give recombinants having altered
CC contents of PUFA (specifically DHA and EPA). Oils from these plants are
CC useful as dietary supplements (in infant feeeding formulations, to give a
CC PUFA profile closer to that of human milk; for treating malnutrition; in
CC intravenous feeding formulations; in cooking oils, fats etc.), also as
CC anti-inflammatory agents and for reducing cholesterol levels. Fragments
CC from the genes are useful as probes to isolate related molecules or to
CC detect organisms that express PKS-like genes. The method facilitates
CC large scale production of PUFA by providing new pathways for their
CC synthesis or suppressing interfering pathways. Expression of PUFA in
CC seeds allows simple recovery, as oil which can be engineered to have a
CC particular PUFA profile. Expression in microbes also allows simple
CC recovery and control of PUFA profile and is not subject to external
CC variables such as weather or food supply. Sequences AAW89396-404
CC represent different ORF proteins of S. putrefaciens PKS-like cluster
XX
SQ Sequence 543 AA;

AAW89403 Length: 543 October 5, 2004 14:15 Type: P Check: 3609 ..

1 MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRAC YVVNHADHGF

51 GIAQTADIVT EQAANSTDLP VSAFTPALGT ESLGDNNFRR VHGVKYAYYA

101 GAMANGISSE ELVIALGQAG ILCGSPGAAG LIPSRVEAAI NRIQAALPNG

151 PYMFNLIHSP SEPALERGSV ELFLKHKVRT VEASAFLGLT PQIVYYRAAG

201 LSRDAQGKVV VGNKVIAKVS RTEVAEKFMM PAPAKMLQKL VDDGSITAEQ

251 MELAQLVPMA DDITAEADSG GHTDNRPLVT LLPTILALKE EIQAKYQYDT

301 PIRVGCGGGV GTPDAALATF NMGAAYIVTG SINQACVEAG ASDHTRKLLA

351 TTEMADVTMA PAADMFEMGV KLQVVKRGTL FPMRANKLYE IYTRYDSIEA

401 IPLDEREKLE KQVFRSSLDE IWAGTVAHFN ERDPKQIERA EGNPKRKMAL

451 IFRWYLGLSS RWSNSGEVGR EMDYQIWAGP ALGAFNQWAK GSYLDNYQDR

501 NAVDLAKHLM YGAAYLNRIN SLTAQGVKVP AQLLRWKPNQ RMA

!!AA_SEQUENCE 1.0
ID **ABM35941** standard; protein; 491 AA.
XX
AC ABM35941;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #617.
XX

KW Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Douglass J;
XX
DR WPI; 2003-381789/36.
DR N-PSDB; ACF64436.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Example 1; SEQ ID NO 617; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 491 AA;

ABM35941 Length: 491 October 5, 2004 14:16 Type: P Check: 2758 ..

1 YVEYGPVAQR IEQQPSNLSR VGSIPTRAAT VSSRVSPGED LAHRKVSRAA

51 GSLSTMSYVA PTPIARRPWW QRALRNPWFW VLLALVLISA ACMVHTYVIV

101 HADTEVEQQG QKGVIPGITN QSLKLAFHYA WPTAAVWSAI FIFLDRFRTR

151 HLHVWFLCFC WGACVATWIS MHVNTWMAGM LSVTGGVDPA SGAGPAVYSA

201 PFVEESCKAL VLFALAIGMG RRMTSVVQTV SMAGLSAIGF AFVENIMYYA

251 RADNYARVTA SAGDPKQAVM ELVLLRGVYA SFGHPLFTSM TGIGLALGLR

301 SRSRLVRIFA PTTGFVMAVV GHMLFNGFSS VLPMAILKKL WFVALGIVAS

351 VVVFLVVRTV REGRMIRYRL EDYVKGGWLP NSDADTLSAL RRRQWAALVA

401 LSQGPRRWWH TLEFLRVGTD LAYLRDAIVR GLDDDPGPRQ IELINRMNVL

451 RPAAITVARG AKLSKPRLFA FLKRRRNQPV NNELQWAPPQ A

!!AA_SEQUENCE 1.0
ID **ABM68937** standard; protein; 591 AA.
XX
AC ABM68937;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #2034.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 2034; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 591 AA;

ABM68937 Length: 591 October 5, 2004 14:16 Type: P Check: 7539 ..

1 VNTMMQTFIP GKDAALEDSI NRFQQKLQHL GFNIEEASWL NPVPHVWSVH

51 IRDRDCPLCF TNGKGASKKA ALASALGEYF ERLSTNYFFA DFYLGEAIAN

```
     101  GDFVHYPNEK WFALPDDNSL PAGILDDRLH QFYNPENELC ASQLVDLQSG

     151  NEERGICSLP FTRQSDHHTV YIPMNIIGNL YVSNGMSAGN TMNEARVQGL

     201  SEVFERYVKN RIIAEGISLP EIPQEVMNRY PDVVAAVSKL QDEGFPLYCY

     251  DASLSGQFPV ICVVLFNPNN GTCFASFGAH PEFGVALERT VTELLQGRSL

     301  KDLDVFNAPT FDNEEVAEHT NLETHFIDSS GLISWDLFKQ DADYAFADWS

     351  FSGTTEEEFA TLMAIFKQLD AEVYIADYEH LGVYACRILV PGMSDIYPVE

     401  DLHLANNTMG IHLREALLAL PDSQWQAEEY LALIEQLDEE GLDDFTRIRE

     451  LLGIAVGKDN GWSNLRIGEL KSMLALAGGD LEQALIWVEW TQDFNTSIFT

     501  AERANYYRCL QTLLLLALEA NKAPAQYYQA FIKMYGQDAV EAASAAISGE

     551  HCFYGLWSVD ADLKSLPAHQ ALLAAYEKLQ KAKCDFWNTK Z

!!AA_SEQUENCE 1.0
ID   ABP96316 standard; protein; 635 AA.
XX
AC   ABP96316;
XX
DT   20-MAY-2003  (first entry)
XX
DE   Caenorhabditis elegans 5-HT3 receptor protein SEQ ID NO:1.
XX
KW   Caenorhabditis elegans; 5-HT3 receptor; helminth; arthropod; ketone;
KW   antihelminthic; pesticide; nematocide; 5-HT3 receptor antagonist.
XX
OS   Caenorhabditis elegans.
XX
PN   WO2003015517-A1.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-AU001096.
XX
PR   14-AUG-2001; 2001AU-00007011.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Trowell SC,  Saubern S,  Liao C;
XX
DR   WPI; 2003-268226/26.
XX
PT   Methods of using substituted ketone derivatives for controlling helminths
PT   and arthropods.
XX
PS   Disclosure; Page 40-43; 53pp; English.
XX
CC   The present invention describes methods for controlling helminths and
CC   arthropods comprising exposing them to a substituted ketone derivative
CC   (I) excluding ondansetron and tropanyl dichlorobenzoate. (I) has
CC   antihelminthic, pesticide and nematocide activities, and can be used as a
CC   5-HT3 receptor antagonist. (I) can be used for controlling helminths and
CC   arthropods, particularly those having a feeding mechanism involving
CC   sucking plant or animal fluids via a muscular pump. The method is
CC   particularly useful for the control of insects and parasitic nematodes.
CC   The compounds modulate the activity of helminth and arthropod 5-HT3
CC   receptors. The present sequence represents a Caenorhabditis elegans 5-HT3
CC   receptor protein, which is given in the exemplification of the present
CC   invention
XX
SQ   Sequence 635 AA;

 ABP96316  Length: 635  October 5, 2004 14:16  Type: P  Check: 9069  ..

       1  MIICYSCLTV SILLTIKFVP CRFAGIEHQN TKSRVHFSLL DSRQENDTNH

      51  FEIAEAKFQK PHNEENTIGT ITKFAPSVQE QHSSAVIPMP HFDQNRLEQA

     101  LRIKGSIDGT EEALYRSLLD HTVYEKDVRP CIHHSQPTNV TFGFLLNQIV

     151  EMDERNQALT TRSWLNINWM DPRLSWNESL WSEIKAIYIP HARIWKPDII

     201  LVNNAIREYY ASLVSTDVMV TSDGNVTWLF SALFRSSCPI RVRYYPFDDQ

     251  QCDLKFASWS HDITEINLGL NTDKGDLSSY MNNSEFDLVD MTAVREVVTF

     301  PSDTNSDWPI IVIRIHMHRR PLFYVFNHIV PCVLISSMAV LGFLMPPETG

     351  EKINMIITTL LSMGVYLQSI TESIPPTSEG VPLIGMYYVS SLLMVCLATC

     401  VNVITLNMHR NGAANQGRHV PAWMQKWILG YLATFMRMSI REPDSIALLK

     451  ASQSKKSTIR RSSILRDLKR VKNMSNVRAK SKEQNANREC ECMDPLVHIY

     501  AESIMSCLAA DTKPMNGSTI REDFASESTF LGRVVSDGIM PRISASSNSV

     551  LTEFETRFRR ILKRVYRSLQ QHEIREEILD ERSRIQWQWQ QLASVVDRLL

     601  LCLFCTATLF TIICLLIVPV AYRDNDSMLS FLNFF

!!AA_SEQUENCE 1.0
ID   ADE78990 standard; protein; 444 AA.
XX
AC   ADE78990;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human protein modification and maintenance molecule (PMMM)-28.
XX
KW   protein modification and maintenance molecule; PMMM;
KW   protein modification; protein maintenance; protein function;
KW   protein conformation; protein stabilisation; protein degradation; kinase;
KW   phosphatase; protease; protease inhibitor; isomerase; transferase;
KW   molecular chaperone; anti-HIV; antiallergic; antiinflammatory;
KW   antianaemic; antiparkinsonian; nootropic; anticonvulsant;
KW   antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
KW   cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;
KW   antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic;
KW   antiparasitic; antihelmintic; antipsoriatic; uropathic; ophthalmological;
KW   antirheumatic; haemostatic; antibacterial; virucide; protozoacide;
KW   fungicide; gene therapy; cell proliferative disorder; arteriosclerosis;
KW   hepatitis; polycythaemia vera; psoriasis; primary thromobocytopaenia;
KW   cancer; developmental disorder; anaemia; mental retardation;
KW   neurological disorder; Alzheimer's disease; Parkinson's disease;
KW   epilepsy; autoimmune disorder; inflammatory disorder; AIDS; allergies;
KW   asthma; autoimmune thyroiditis; Crohn's disease; diabetes mellitus;
KW   glomerulonephritis; Goodpasture's syndrome; multiple sclerosis;
KW   arthritis; osteoporosis; pancreatitis; Sjogren's syndrome;
KW   microbial infection; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 30
FT                   /note= "May be substituted by Glu as a result of a single
FT                   nucleotide polymorphism"
FT   Misc-difference 235
FT                   /note= "May be substituted by His as a result of a single
FT                   nucleotide polymorphism"
XX
PN   WO2003063688-A2.
XX
PD   07-AUG-2003.
XX
PF   23-JAN-2003; 2003WO-US002500.
XX
PR   25-JAN-2002; 2002US-0351928P.
PR   25-FEB-2002; 2002US-0359903P.
PR   21-MAR-2002; 2002US-0366837P.
```

```
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hafalia AJA, Li JX, Gorvad AE, Chawla NK, Sprague WW, Lee SY;
PI   Chang H, Elliott VS, Ramkumar J, Khare R, Emerling BM, Kable AE;
PI   Tang YT, Yue H, Gietzen KJ, Lee S, Swarnakar A, Baughn MR;
PI   Wilson AD, Jin P, Chien D, Hawkins PR, Jiang X, Jackson AA;
PI   Bhatia U, Burrill JD, Blake JJ, Ho A, Zheng W, Ison CH, Marquis JP;
PI   Tran UK, Lal PG, Warren BA, Xu Y, Honchell CD, Becha SD;
PI   Lehr-Mason PM;
XX
DR   WPI; 2003-636761/60.
DR   N-PSDB; ADE79048.
XX
PT   New human protein modification and maintenance molecules and
PT   polynucleotides, useful for diagnosing, treating or preventing autoimmune
PT   or inflammatory disorders (e.g. AIDS, allergy or anemia), multiple
PT   sclerosis or cancer.
XX
PS   Claim 1; SEQ ID NO 28; 405pp; English.
XX
CC   This invention relates to novel isolated human proteins, which are human
CC   protein modification and maintenance molecules (PMMM). The cellular
CC   processes regulating modification and maintenance of protein molecules
CC   coordinate their function, conformation, stabilisation and degradation.
CC   Each of these processes is mediated by key enzymes or proteins such as
CC   kinases, phosphatases, proteases, protease inhibitors, isomerases,
CC   transferases and molecular chaperones. Compounds which modulate the
CC   proteins of the invention may have anti-HIV, antiallergic,
CC   antiinflammatory, antianaemic, antiparkinsonian, nootropic,
CC   anticonvulsant, antiarteriosclerotic, antiasthmatic, immunosuppressive,
CC   antithyroid, cytostatic, hepatotropic, dermatological, antidiabetic,
CC   nephrotropic, antigout, thyromimetic, neuroprotective, osteopathic,
CC   antiarthritic, antiparasitic, antihelmintic, antipsoriatic, uropathic,
CC   ophthalmological, antirheumatic, haemostatic, antibacterial, virucide,
CC   protozoacide or fungicide activities. The DNA sequence which encode the
CC   proteins of the invention may be useful for gene therapy. The human
CC   protein modification and maintenance molecules (PMMMs), the DNA sequences
CC   which encode them and their modulating compounds are useful for
CC   diagnosing, treating or preventing disorders associated with aberrant
CC   expression of PMMM, particularly cell proliferative disorders (for
CC   example arteriosclerosis, hepatitis, polycythaemia vera, psoriasis,
CC   primary thromobocytopaenia or cancer), developmental disorders (for
CC   example anaemia or mental retardation), neurological disorders (for
CC   example Alzheimer's disease, Parkinson's disease or epilepsy),
CC   autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
CC   autoimmune thyroiditis, Crohn's disease, diabetes mellitus,
CC   glomerulonephritis, Goodpasture's syndrome, multiple sclerosis,
CC   arthritis, osteoporosis, pancreatitis, Sjogren's syndrome) or microbial
CC   infections. The present sequence is the amino acid sequence of a human
CC   PMMM of the invention.
XX
SQ   Sequence 444 AA;

 ADE78990  Length: 444  October 5, 2004 14:17  Type: P  Check: 5455  ..

       1  MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

      51  GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

     101  CKDGLGEYTC TCLEGFEGKN CELWPYPCGK QTLERRKRSV AQATSSSGEA

     151  PDSITWKPYD AADLDPTENP FDLLDFNQTQ PERGDNNLTR IVGGQECKDG

     201  ECPWQALLIN EENEGFCGGT ILSEFYILTA AHCLYQAKRF KVRVGDRNTE

     251  QEEGGEAVHE VEVVIKHNRF TKETYDFDIA VLRLKTPITF RMNVAPACLP

     301  ERDWAESTLM TQKTGIVSGF GRTHEKGRQS TRLKMLEVPY VDRNSCKLSS

     351  SFIITQNMFC AGYDTKQEDA CQGDSGGPHV TRFKDTYFVT GIVSWGEGCA

     401  RKGKYGIYTK VTAFLKWIDR SMKTRGLPKA KSHAPEVITS SPLK
```

```
!!AA_SEQUENCE 1.0
ID   ADE78991 standard; protein; 377 AA.
XX
AC   ADE78991;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human protein modification and maintenance molecule (PMMM)-29.
XX
KW   protein modification and maintenance molecule; PMMM;
KW   protein modification; protein maintenance; protein function;
KW   protein conformation; protein stabilisation; protein degradation; kinase;
KW   phosphatase; protease; protease inhibitor; isomerase; transferase;
KW   molecular chaperone; anti-HIV; antiallergic; antiinflammatory;
KW   antianaemic; antiparkinsonian; nootropic; anticonvulsant;
KW   antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
KW   cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;
KW   antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic;
KW   antiparasitic; antihelmintic; antipsoriatic; uropathic; ophthalmological;
KW   antirheumatic; haemostatic; antibacterial; virucide; protozoacide;
KW   fungicide; gene therapy; cell proliferative disorder; arteriosclerosis;
KW   hepatitis; polycythaemia vera; psoriasis; primary thromobocytopaenia;
KW   cancer; developmental disorder; anaemia; mental retardation;
KW   neurological disorder; Alzheimer's disease; Parkinson's disease;
KW   epilepsy; autoimmune disorder; inflammatory disorder; AIDS; allergies;
KW   asthma; autoimmune thyroiditis; Crohn's disease; diabetes mellitus;
KW   glomerulonephritis; Goodpasture's syndrome; multiple sclerosis;
KW   arthritis; osteoporosis; pancreatitis; Sjogren's syndrome;
KW   microbial infection; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 30
FT                   /note= "May be substituted by Glu as a result of a single
FT                   nucleotide polymorphism"
XX
PN   WO2003063688-A2.
XX
PD   07-AUG-2003.
XX
PF   23-JAN-2003; 2003WO-US002500.
XX
PR   25-JAN-2002; 2002US-0351928P.
PR   25-FEB-2002; 2002US-0359903P.
PR   21-MAR-2002; 2002US-0366837P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hafalia AJA, Li JX, Gorvad AE, Chawla NK, Sprague WW, Lee SY;
PI   Chang H, Elliott VS, Ramkumar J, Khare R, Emerling BM, Kable AE;
PI   Tang YT, Yue H, Gietzen KJ, Lee S, Swarnakar A, Baughn MR;
PI   Wilson AD, Jin P, Chien D, Hawkins PR, Jiang X, Jackson AA;
PI   Bhatia U, Burrill JD, Blake JJ, Ho A, Zheng W, Ison CH, Marquis JP;
PI   Tran UK, Lal PG, Warren BA, Xu Y, Honchell CD, Becha SD;
PI   Lehr-Mason PM;
XX
DR   WPI; 2003-636761/60.
DR   N-PSDB; ADE79049.
XX
PT   New human protein modification and maintenance molecules and
PT   polynucleotides, useful for diagnosing, treating or preventing autoimmune
PT   or inflammatory disorders (e.g. AIDS, allergy or anemia), multiple
PT   sclerosis or cancer.
XX
PS   Claim 1; SEQ ID NO 29; 405pp; English.
XX
CC   This invention relates to novel isolated human proteins, which are human
CC   protein modification and maintenance molecules (PMMM). The cellular
CC   processes regulating modification and maintenance of protein molecules
CC   coordinate their function, conformation, stabilisation and degradation.
CC   Each of these processes is mediated by key enzymes or proteins such as
```

CC kinases, phosphatases, proteases, protease inhibitors, isomerases,
CC transferases and molecular chaperones. Compounds which modulate the
CC proteins of the invention may have anti-HIV, antiallergic,
CC antiinflammatory, antianaemic, antiparkinsonian, nootropic,
CC anticonvulsant, antiarteriosclerotic, antiasthmatic, immunosuppressive,
CC antithyroid, cytostatic, hepatotropic, dermatological, antidiabetic,
CC nephrotropic, antigout, thyromimetic, neuroprotective, osteopathic,
CC antiarthritic, antiparasitic, antihelmintic, antipsoriatic, uropathic,
CC ophthalmological, antirheumatic, haemostatic, antibacterial, virucide,
CC protozoacide or fungicide activities. The DNA sequence which encode the
CC proteins of the invention may be useful for gene therapy. The human
CC protein modification and maintenance molecules (PMMMs), the DNA sequences
CC which encode them and their modulating compounds are useful for
CC diagnosing, treating or preventing disorders associated with aberrant
CC expression of PMMM, particularly cell proliferative disorders (for
CC example arteriosclerosis, hepatitis, polycythaemia vera, psoriasis,
CC primary thromobocytopaenia or cancer), developmental disorders (for
CC example anaemia or mental retardation), neurological disorders (for
CC example Alzheimer's disease, Parkinson's disease or epilepsy),
CC autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
CC autoimmune thyroiditis, Crohn's disease, diabetes mellitus,
CC glomerulonephritis, Goodpasture's syndrome, multiple sclerosis,
CC arthritis, osteoporosis, pancreatitis, Sjogren's syndrome) or microbial
CC infections. The present sequence is the amino acid sequence of a human
CC PMMM of the invention.
XX
SQ Sequence 377 AA;

ADE78991 Length: 377 October 5, 2004 14:17 Type: P Check: 9890 ..

```
       1  MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

      51  GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

     101  CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS

     151  CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW

     201  KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA

     251  LLINEENEKG RQSTRLKMLE VPYVDRNSCK LSSSFIITQN MFCAGYDTKQ

     301  EDACQGDSGG PHVTRFKDTY FVTGIVSWGE GCARKGKYGI YTKVTAFLKW

     351  IDRSMKTRGL PKAKSHAPEV ITSSPLK
```

!!AA_SEQUENCE 1.0
ID **ADE78992** standard; protein; 442 AA.
XX
AC ADE78992;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human protein modification and maintenance molecule (PMMM)-30.
XX
KW protein modification and maintenance molecule; PMMM;
KW protein modification; protein maintenance; protein function;
KW protein conformation; protein stabilisation; protein degradation; kinase;
KW phosphatase; protease; protease inhibitor; isomerase; transferase;
KW molecular chaperone; anti-HIV; antiallergic; antiinflammatory;
KW antianaemic; antiparkinsonian; nootropic; anticonvulsant;
KW antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
KW cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;
KW antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic;
KW antiparasitic; antihelmintic; antipsoriatic; uropathic; ophthalmological;
KW antirheumatic; haemostatic; antibacterial; virucide; protozoacide;
KW fungicide; gene therapy; cell proliferative disorder; arteriosclerosis;
KW hepatitis; polycythaemia vera; psoriasis; primary thromobocytopaenia;
KW cancer; developmental disorder; anaemia; mental retardation;
KW neurological disorder; Alzheimer's disease; Parkinson's disease;
KW epilepsy; autoimmune disorder; inflammatory disorder; AIDS; allergies;
KW asthma; autoimmune thyroiditis; Crohn's disease; diabetes mellitus;
KW glomerulonephritis; Goodpasture's syndrome; multiple sclerosis;
KW arthritis; osteoporosis; pancreatitis; Sjogren's syndrome;
KW microbial infection; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 30
FT /note= "May be substituted by Glu as a result of a single
FT nucleotide polymorphism"
FT Misc-difference 354
FT /note= "May be substituted by a stop codon as a result of
FT a single nucleotide polymorphism"
XX
PN WO2003063688-A2.
XX
PD 07-AUG-2003.
XX
PF 23-JAN-2003; 2003WO-US002500.
XX
PR 25-JAN-2002; 2002US-0351928P.
PR 25-FEB-2002; 2002US-0359903P.
PR 21-MAR-2002; 2002US-0366837P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hafalia AJA, Li JX, Gorvad AE, Chawla NK, Sprague WW, Lee SY;
PI Chang H, Elliott VS, Ramkumar J, Khare R, Emerling BM, Kable AE;
PI Tang YT, Yue H, Gietzen KJ, Lee S, Swarnakar A, Baughn MR;
PI Wilson AD, Jin P, Chien D, Hawkins PR, Jiang X, Jackson AA;
PI Bhatia U, Burrill JD, Blake JJ, Ho A, Zheng W, Ison CH, Marquis JP;
PI Tran UK, Lal PG, Warren BA, Xu Y, Honchell CD, Becha SD;
PI Lehr-Mason PM;
XX
DR WPI; 2003-636761/60.
DR N-PSDB; ADE79050.
XX
PT New human protein modification and maintenance molecules and
PT polynucleotides, useful for diagnosing, treating or preventing autoimmune
PT or inflammatory disorders (e.g. AIDS, allergy or anemia), multiple
PT sclerosis or cancer.
XX
PS Claim 1; SEQ ID NO 30; 405pp; English.
XX
CC This invention relates to novel isolated human proteins, which are human
CC protein modification and maintenance molecules (PMMM). The cellular
CC processes regulating modification and maintenance of protein molecules
CC coordinate their function, conformation, stabilisation and degradation.
CC Each of these processes is mediated by key enzymes or proteins such as
CC kinases, phosphatases, proteases, protease inhibitors, isomerases,
CC transferases and molecular chaperones. Compounds which modulate the
CC proteins of the invention may have anti-HIV, antiallergic,
CC antiinflammatory, antianaemic, antiparkinsonian, nootropic,
CC anticonvulsant, antiarteriosclerotic, antiasthmatic, immunosuppressive,
CC antithyroid, cytostatic, hepatotropic, dermatological, antidiabetic,
CC nephrotropic, antigout, thyromimetic, neuroprotective, osteopathic,
CC antiarthritic, antiparasitic, antihelmintic, antipsoriatic, uropathic,
CC ophthalmological, antirheumatic, haemostatic, antibacterial, virucide,
CC protozoacide or fungicide activities. The DNA sequence which encode the
CC proteins of the invention may be useful for gene therapy. The human
CC protein modification and maintenance molecules (PMMMs), the DNA sequences
CC which encode them and their modulating compounds are useful for
CC diagnosing, treating or preventing disorders associated with aberrant
CC expression of PMMM, particularly cell proliferative disorders (for
CC example arteriosclerosis, hepatitis, polycythaemia vera, psoriasis,
CC primary thromobocytopaenia or cancer), developmental disorders (for
CC example anaemia or mental retardation), neurological disorders (for
CC example Alzheimer's disease, Parkinson's disease or epilepsy),
CC autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
CC autoimmune thyroiditis, Crohn's disease, diabetes mellitus,
CC glomerulonephritis, Goodpasture's syndrome, multiple sclerosis,
CC arthritis, osteoporosis, pancreatitis, Sjogren's syndrome) or microbial
CC infections. The present sequence is the amino acid sequence of a human
CC PMMM of the invention.

XX
SQ Sequence 442 AA;

ADE78992 Length: 442 October 5, 2004 14:17 Type: P Check: 53 ..

```
     1  MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK
    51  GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK
   101  CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS
   151  CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW
   201  KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
   251  LLINEENEVE VVIKHNRFTK ETYDFDIAVL RLKTPITFRM NVAPACLPER
   301  DWAESTLMTQ KTGIVSGFGR THEKGRQSTR LKMLEVPYVD RNSCKLSSSF
   351  IITQNMFCAG YDTKQEDACQ GDSGGPHVTR FKDTYFVTGI VSWGEGCARK
   401  GKYGIYTKVT AFLKWIDRSM KTRGLPKAKS HAPEVITSSP LK
```

PIR
Citations

!!AA_SEQUENCE 1.0
P1;A89606 - protein F18G5.4 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
C;Accession: A89606
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology.
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_elegans/ for a list of authors
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 285, 1493, 1999
A;Accession: A89606
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-692 <STO>
A;Cross-references: GB:chr_X; PIDN:AAA81080.1; PID:g1055091; GSPDB:GN00028; CESP:F18G5.4
C;Genetics:
A;Gene: F18G5.4
A;Map position: X

A89606 Length: 692 October 5, 2004 14:20 Type: P Check: 9797 ..

1 MKHVASLAHC FFHLGRRCLK LRPHKLEEPR HRSKLVRRAL LPLLLHSFAA

51 FFFLFCAPFH TITFALHITY VRREDLLLLL PLCICCAPAH HPRSFSNWLL

101 KQSVQSTGPQ FPPKFQKPHN EENTIGTITK FAPSVQEQHS SAVIPMPHFD

151 QNRLEQALRI KGSIDGTEEA LYRSLLDHTV YEKDVRPCIH HSQPTNVTFG

201 FLLNQIVEMD ERNQALTTRS WLNINWMDPR LSWNESLWSE IKAIYIPHAR

251 IWKPDIILVN NRQTKNSDAC LHAYISASEP AAIREYYASL VSTDVMVTSD

301 GNVTWLFSAL FRSSCPIRVR YYPFDDQQCD LKFASWSHDI TEINLGLNTD

351 KGDLSSYMNN SEFDLVDMTA VREVVTFPSD TNSDWPIIVI RIHMHRRPLF

401 YVFNHIVPCV LISSMAVLGF LMPPETGEKI NMIITTLLSM GVYLQSITES

451 IPPTSEGVPL IGMYYVSSLL MVCLATCVNV ITLNMHRNGA ANQGRHVPAW

501 MQKWILGYLA TFMRMSIREP DSIALLKASQ SKKSTIRRSS ILRDLKRVKN

551 MSNVRAKSKE QNANRECECM DPLVHIYAES IMSCLAADTK PMNGSTIRED

601 FASESTFLGR VVSDGIMPRI SASSNSVLTE FETRFRRILK RVYRSLQQHE

651 IREEILDERS RIQCSGNNLH LSLIDFYYVF FALQHCSQSS AS

!!AA_SEQUENCE 1.0
P1;AE0613 - conserved hypothetical protein STY0975 [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain CT18)
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AE0613
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Mungall, K.L.; Bentley, S.D.; Holden, M.T.G.; Sebaihia, M.; Baker, S.; Basham, D.; Brooks, K.; Chillingworth, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, J.; Feltwell, T.; Hamlin, N.; Haque, A.; Hien, T.T.; Holroyd, S.; Jagels, K.; Krogh, A.; Larsen, T.S.; Leather, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.G.
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AE0613
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-589 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05375.1; PID:g16502138; GSPDB:GN00176
C;Genetics:
A;Gene: STY0975

AE0613 Length: 589 October 5, 2004 14:20 Type: P Check: 3304 ..

1 MIFMTQTFIP GKDAALEDSI ARFQQKLLDL GFHIEEASWL NPVPNVWSVH

51 IRDKECALCF TNGKGATKKA ALASALGEYF ERLSTNYFFA DFWLGETIAN

101 GPFVHYPNEK WFSLTENDDV PEGLLDARLR AFYDPENELT GSQLIDLQSG

151 NEERGVCGLP FTRQSDNQTV YIPMNIIGNL YVSNGMSAGN TRNEARVQGL

201 SEVFERYVKN RIIAESISLP EIPAEVMARY PAVMESIATL EAEGFPIFAY

251 DGSLGGKYPV ICVVLFNPAN GTCFASFGAH PDFGVALERT VTELLQGRGL

301 KDLDVFTPPT FDDEEVAEHT NLETHFIDSS GLISWDLFKQ DADYPFTDWS

351 FSGTTEEEFA TLMAIFAAED KEVYIADYEH LGVYACRIIV PGMSDIYPAE

401 DLWLANNNMG SHLREILLSL PGSAWNKEDY LNLIEQLDEE GFDDFTRVRE

451 LLGLATGADN GWYTLRVGEL KAMLALAGGD LEQALIWTEW TMEFNSSVFS

501 PVRANYYRCL QTLLLLSQED ARQPLQYLNA FIKMYGAEAV EAASTALSGE

551 AAFYGLPAVD HDLQAFPAHQ SLLKAYDKLQ RAKAAYWSK

!!AA_SEQUENCE 1.0
P1;B90396 - hypothetical protein SSO2259 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: B90396
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Weiher, C.C.Y.; Clausen, I.G.; Curtis, B.A.; De Moors, A.; Erauso, G.; Fletcher, C.; Gordon, P.M.K.; Heikamp-de Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.; Schenk, M.E.; Theriault, C.; Tolstrup, N.; Charlebois, R.L.; Doolittle, W.F.; Duguet, M.; Gaasterland, T.; Garrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: B90396
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-220 <KUR>
A;Cross-references: GB:AE006641; NID:g13815561; PIDN:AAK42425.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO2259

B90396 Length: 220 October 5, 2004 14:20 Type: P Check: 6793 ..

1 MTNGEFYALD NLLKDEKLNS LNKILDVINT TDRLGILDVI KGILEDENTI

51 GKIIGSLTSD DVLELLVNWD KVIKTTKLFI NEDNIYNIQF LINLIDKVRS

101 KGILDPIIGL LEDEESLGKI INALINDFTL NLINHWEEII NDLSRIDLTN

151 FKYYTLLVSA TGEALKTENV KPITSIWEIY KLLKDPDIQR GLGVAASVLK

201 RIGKLYVPDK GLAFEVEKKL

!!AA_SEQUENCE 1.0
P1;EXHU - coagulation factor Xa (EC 3.4.21.6) precursor [validated] - human
N;Alternate names: Stuart factor
C;Species: Homo sapiens (man)
C;Date: 15-Nov-1984 #sequence_revision 02-May-1994 #text_change 08-Dec-2000

C;Accession: A24478; JQ0917; A42485; A25853; A22208; A21284; A20362; S39415; I54051; A00924
R;Leytus, S.P.; Foster, D.C.; Kurachi, K.; Davie, E.W.
Biochemistry 25, 5098-5102, 1986
A;Title: Gene for human Factor X: a blood coagulation factor whose gene organization is essentially identical with that of Factor IX and protein C.
A;Reference number: A24478; MUID:87026600; PMID:3768336
A;Accession: A24478
A;Molecule type: DNA
A;Residues: 1-488 <LEY>
A;Cross-references: GB:L29433; GB:M14327; NID:g459809; PIDN:AAA52764.1; PID:g182831
R;Messier, T.L.; Pittman, D.D.; Long, G.L.; Kaufman, R.J.; Church, W.R.
Gene 99, 291-294, 1991
A;Title: Cloning and expression in COS-1 cells of a full-length cDNA encoding human coagulation factor X.
A;Reference number: JQ0917; MUID:91216473; PMID:1902434
A;Accession: JQ0917
A;Molecule type: mRNA
A;Residues: 1-488 <MES>
A;Cross-references: GB:M57285; NID:g182389; PIDN:AAA52421.1; PID:g182390
R;Miao, C.H.; Leytus, S.P.; Chung, D.W.; Davie, E.W.
J. Biol. Chem. 267, 7395-7401, 1992
A;Title: Liver-specific expression of the gene coding for human factor X, a blood coagulation factor.
A;Reference number: A42485; MUID:92218390; PMID:1313796
A;Accession: A42485
A;Molecule type: DNA
A;Residues: 1-15 <MIA>
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIN:93780, NCBIP:93787)
R;Kaul, R.K.; Hildebrand, B.; Roberts, S.; Jagadeeswaran, P.
Gene 41, 311-314, 1986
A;Title: Isolation and characterization of human blood-coagulation factor X cDNA.
A;Reference number: A25853; MUID:86221713; PMID:3011603
A;Accession: A25853
A;Molecule type: mRNA
A;Residues: 19-284,'E',289-488 <KAU>
A;Cross-references: GB:M22613; NID:g180335; PIDN:AAA51984.1; PID:g180336
R;Fung, M.R.; Hay, C.W.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 3591-3595, 1985
A;Title: Characterization of an almost full-length cDNA coding for human blood coagulation factor X.
A;Reference number: A22208; MUID:85216545; PMID:2582420
A;Accession: A22208
A;Molecule type: mRNA
A;Residues: 13-441,'S',443-488 <FUN>
A;Cross-references: GB:K03194; NID:g182840; PIDN:AAA52490.1; PID:g182841
R;Leytus, S.P.; Chung, D.W.; Kisiel, W.; Kurachi, K.; Davie, E.W.
Proc. Natl. Acad, Sci. U.S.A. 81, 3699-3702, 1984
A;Title: Characterization of a cDNA coding for human factor X.
A;Reference number: A21284; MUID:84222026; PMID:6587384
A;Accession: A21284
A;Molecule type: mRNA
A;Residues: 13-284,'E',289-488 <LE2>
A;Cross-references: GB:K01886
R;McMullen, B.A.; Fujikawa, K.; Kisiel, W.; Sasagawa, T.; Howald, W.N.; Kwa, E.Y.; Weinstein, B.
Biochemistry 22, 2875-2884, 1983
A;Title: Complete amino acid sequence of the light chain of human blood coagulation factor X: evidence for identification of residue 63 as beta-hydroxyaspartic acid.
A;Reference number: A20362; MUID:83257207; PMID:6871167
A;Accession: A20362
A;Molecule type: protein
A;Residues: 41-179 <MCM>
R;Inoue, K.; Morita, T.
Eur. J. Biochem. 218, 153-163, 1993
A;Title: Identification of O-linked oligosaccharide chains in the activation peptides of blood coagulation factor X. The role of the carbohydrate moieties in the activation of factor X.
A;Reference number: S39414; MUID:94062825; PMID:8243461
A;Accession: S39415
A;Molecule type: protein
A;Residues: 183-234 <INO>
A;Note: glycosylation sites
A;Note: identification and characterization of beta-hydroxyaspartic acid
R;Jagadeeswaran, P.; Reddy, S.V.; Rao, K.J.; Hamsabhushanam, K.; Lyman, G.
Gene 84, 517-519, 1989
A;Title: Cloning and characterization of the 5' end (exon 1) of the gene encoding human factor X.
A;Reference number: I54051; MUID:90128299; PMID:2612918
A;Accession: I54051
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-23 <RES>
A;Cross-references: GB:M33297; NID:g183860; PIDN:AAA52636.1; PID:g553330
R;Padmanabhan, K.; Padmanabhan, K.P.; Tulinsky, A.; Park, C.H.; Bode, W.; Huber, R.; Blankenship, D.T.; Cardin, A.D.; Kisiel, W.
J. Mol. Biol. 232, 947-966, 1993
A;Title: Structure of human des(1-45) factor Xa at 2.2 angstroms resolution.
A;Reference number: A49458; MUID:93360277; PMID:8355279
A;Contents: annotation; X-ray crystallography, 2.2 angstroms
C;Comment: The two chains held together by one disulfide bond are formed from a single-chain precursor by excision of residues 180-182.
C;Comment: The activation peptide is cleaved by factor IXa (in the intrinsic pathway) or by factor VIIa (in the extrinsic pathway).
C;Genetics:
A;Gene: GDB:F10
A;Cross-references: GDB:119890; OMIM:227600
A;Map position: 13q34-13q34
A;Introns: 24/1; 77/3; 86/1; 124/1; 150/3; 249/3; 289/1
A;Note: deficiency of this factor causes Stuart disease
C;Function:
A;Description: catalyzes the proteolytic activation of prothrombin to thrombin in the presence of calcium, phospholipid, and factor Va
A;Pathway: blood coagulation
C;Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology
C;Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglutamic acid; glycoprotein; hydrolase; plasma; serine proteinase; vitamin K
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-40/Domain: propeptide #status predicted <PRO>
F;25-84/Domain: Gla domain homology <GLA>
F;41-179/Product: coagulation factor X light chain #status experimental <LCH>
F;90-121/Domain: EGF homology <EG1>
F;129-164/Domain: EGF homology <EG2>
F;183-488/Product: coagulation factor X heavy chain #status experimental <HCH>
F;183-234/Domain: activation peptide #status experimental <APT>
F;235-488/Product: coagulation factor Xa heavy chain #status experimental <ACT>
F;235-462/Domain: trypsin homology <TRY>
F;46,47,54,56,59,60,65,66,69,72,79/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental
F;57-62/Disulfide bonds: #status predicted
F;90-101,95-110,112-121,129-140,136-149,151-164,172-342,241-246,261-277,390-404, 415-443/Disulfide bonds: #status experimental
F;103/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
F;199,211/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;221,231/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;234-235/Cleavage site: Arg-Ile (coagulation factor IXa, coagulation factor VIIa) #status experimental
F;276,322,419/Active site: His, Asp, Ser #status experimental

EXHU Length: 488 October 5, 2004 14:19 Type: P Check: 2459 ..

1 MGRPLHLVLL SASLAGLLLL GESLFIRREQ ANNILARVTR ANSFLEEMKK

51 GHLERECMEE TCSYEEAREV FEDSDKTNEF WNKYKDGDQC ETSPCQNQGK

101 CKDGLGEYTC TCLEGFEGKN CELFTRKLCS LDNGDCDQFC HEEQNSVVCS

151 CARGYTLADN GKACIPTGPY PCGKQTLERR KRSVAQATSS SGEAPDSITW

```
     201  KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA

     251  LLINEENEGF CGGTILSEFY ILTAAHCLYQ AKRFKVRVGD RNTEQEEGGE

     301  AVHEVEVVIK HNRFTKETYD FDIAVLRLKT PITFRMNVAP ACLPERDWAE

     351  STLMTQKTGI VSGFGRTHEK GRQSTRLKML EVPYVDRNSC KLSSSFIITQ

     401  NMFCAGYDTK QEDACQGDSG GPHVTRFKDT YFVTGIVSWG EGCARKGKYG

     451  IYTKVTAFLK WIDRSMKTRG LPKAKSHAPE VITSSPLK

!!AA_SEQUENCE 1.0
P1;H64829 - ycaO protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: H64829
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
 Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
 J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H64829
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-589 <BLAT>
A;Cross-references: GB:AE000192; GB:U00096; NID:g1787125; PIDN:AAC73991.1;
 PID:g1787133; UWGP:b0905
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: ycaO

  H64829  Length: 589  October 5, 2004 14:19  Type: P  Check: 2176  ..

       1  MIFMTQTFIP GKDAALEDSI ARFQQKLSDL GFQIEEASWL NPVPNVWSVH

      51  IRDKECALCF TNGKGATKKA ALASALGEYF ERLSTNYFFA DFWLGETIAN

     101  GPFVHYPNEK WFPLTENDDV PEGLLDDRLR AFYDPENELT GSMLIDLQSG

     151  NEDRGICGLP FTRQSDNQTV YIPMNIIGNL YVSNGMSAGN TRNEARVQGL

     201  SEVFERYVKN RIIAESISLP EIPADVLARY PAVVEAIETL EAEGFPIFAY

     251  DGSLGGQYPV ICVVLFNPAN GTCFASFGAH PDFGVALERT VTELLQGRGL

     301  KDLDVFTPPT FDDEEVAEHT NLETHFIDSS GLISWDLFKQ DADYPFVDWN

     351  FSGTTEEEFA TLMAIFNKED KEVYIADYEH LGVYACRIIV PGMSDIYPAE

     401  DLWLANNSMG SHLRETILSL PGSEWEKEDY LNLIEQLDEE GFDDFTRVRE

     451  LLGLATGSDN GWYTLRIGEL KAMLALAGGD LEQALVWTEW TMEFNSSVFS

     501  PERANYYRCL QTLLLLAQEE DRQPLQYLNA FVRMYGADAV EAASAAMSGE

     551  AAFYGLQPVD SDLHAFAAHQ SLLKAYEKLQ RAKAAFWAK

!!AA_SEQUENCE 1.0
P1;S08438 - vpx protein - human immunodeficiency virus type 2 D205
C;Species: human immunodeficiency virus type 2 D205, HIV-2 D205
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 27-Jan-1995
C;Accession: S08438
R;Dietrich, U.; Adamski, M.; Kreutz, R.; Seipp, A.; Kuehnel, H.;
 Ruebsamen-Waigmann, H.
Nature 342, 948-950, 1989
A;Title: A highly divergent HIV-2-related isolate.
A;Reference number: S08434; MUID:90081881; PMID:2594088
A;Accession: S08438
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-111 <DIE>
A;Cross-references: EMBL:X16109
A;Note: this sequence was submitted to the EMBL Data Library, Aug-1989
C;Genetics:
A;Gene: vpx
C;Superfamily: AIDS vpu protein

  S08438  Length: 111  October 5, 2004 14:19  Type: P  Check: 2803  ..

       1  MDPRERVPPG NSDEETVGEA FAWLERTITE LNRVAVNHLP RELIFQVWQR

      51  SWAYWREEQG MSISYTKYRY LLLMQKAMFV HYTKGCRCLQ EGHGPGGWRS

     101  GPPPPPPPGL A

!!AA_SEQUENCE 1.0
P1;T30186 - hypothetical protein 8 - Shewanella sp.
C;Species: Shewanella sp.
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T30186
R;Takeyama, H.; Takeda, D.; Yazawa, K.; Yamada, A.; Matsunaga, T.
Microbiology 143, 2725-2731, 1997
A;Title: Expression of the eicosapentaenoic acid synthesis gene cluster from
 Shewanella sp. in a transgenic marine cyanobacterium, Synechococcus sp.
A;Reference number: Z20764; MUID:97419510; PMID:9274025
A;Accession: T30186
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-543 <TAK>
A;Cross-references: EMBL:U73935; NID:g2529413; PID:g2529421; PIDN:AAB81126.1
A;Experimental source: strain SCRC-2738

  T30186  Length: 543  October 5, 2004 14:20  Type: P  Check: 3609  ..

       1  MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRAC YVVNHADHGF

      51  GIAQTADIVT EQAANSTDLP VSAFTPALGT ESLGDNNFRR VHGVKYAYYA

     101  GAMANGISSE ELVIALGQAG ILCGSFGAAG LIPSRVEAAI NRIQAALPNG

     151  PYMFNLIHSP SEPALERGSV ELFLKHKVRT VEASAFLGLT PQIVYYRAAG

     201  LSRDAQGKVV VGNKVIAKVS RTEVAEKFMM PAPAKMLQKL VDDGSITAEQ

     251  MELAQLVPMA DDITAEADSG GHTDNRPLVT LLPTILALKE EIQAKYQYDT

     301  PIRVGCGGGV GTPDAALATF NMGAAYIVTG SINQACVEAG ASDHTRKLLA

     351  TTEMADVTMA PAADMFEMGV KLQVVKRGTL FPMRANKLYE IYTRYDSIEA

     401  IPLDEREKLE KQVFRSSLDE IWAGTVAHFN ERDPKQIERA EGNPKRKMAL

     451  IFRWYLGLSS RWSNSGEVGR EMDYQIWAGP ALGAFNQWAK GSYLDNYQDR

     501  NAVDLAKHLM YGAAYLNRIN SLTAQGVKVP AQLLRWKPNQ RMA

!!AA_SEQUENCE 1.0
P1;WZBE17 - gene 17 protein - human herpesvirus 3
C;Species: human herpesvirus 3, varicella-zoster virus
C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C;Accession: H27342
R;Davison, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A;Title: The complete DNA sequence of varicella-zoster virus.
A;Reference number: A27345; MUID:86306657; PMID:3018124
A;Accession: H27342
A;Molecule type: DNA
A;Residues: 1-455 <DAV>
A;Cross-references: EMBL:X04370; NID:g59989; PIDN:CAA27900.1; PID:g60006
C;Genetics:
A;Gene: 17
C;Superfamily: varicella-zoster virus gene 17 protein

  WZBE17  Length: 455  October 5, 2004 14:19  Type: P  Check: 3936  ..
```

```
     1  MGLFGLTRFI HEHKLVKPSI ISTPPGVLTP VAVDVWNVMY TLLERLYPVG
    51  KRENLHGPSV TIHCLGVLLR LLTQRSYYPI FVLERCTDGP LSRGAKAIMS
   101  RAMNHDERGT SDLTRVLLSS NTSCSIKYNK TSETYDSVFR NSSTSCIPSE
   151  ENKSQDMFLD GCPRQTDKTI CLRDQNVCSL TSTMPSRGHP NHRLYHKLCA
   201  SLIRWMGYAY VEAVDIEADE ACANLFHTRT VALVYTTDTD LLFMGCDILL
   251  DAIPMFAPVV RCRDLLQYLG ITYPEFLVAF VRCQTDLHTS DNLKSVQQVI
   301  QDTGLKVPHQ MDTSTRSPTY DSWRHGEVFK SLTVATSGKT ENGVSVSKYA
   351  SNRSEVTVDA SWALNLLPPS SSPLDNLERA FVEHIIAVVT PLTRGRLKLM
   401  KRVNIMQNTA DPYMVINTLY HNLKGEKMAR QYARIFKQFI PTPLPLNTVL
   451  TKYWN

!!AA_SEQUENCE 1.0
P1;XPBE12 - major antigenic structural protein p100 - human herpesvirus 6
 (strain U1102)
C;Species: human herpesvirus 6
C;Date: 30-Jun-1993 #sequence_revision 29-Oct-1999 #text_change 21-Jul-2000
C;Accession: T09303; A42533
R;Nicholas, J.; Martin, M.
J. Virol. 68, 597-610, 1994
A;Title: Nucleotide sequence analysis of a 38.5-kilobase-pair region of the
 genome of human herpesvirus 6 encoding human cytomegalovirus immediate-early
 gene homologs and transactivating functions.
A;Reference number: Z16644; MUID:94118404; PMID:8289364
A;Accession: T09303
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-871 <NIC>
A;Cross-references: EMBL:L25528; NID:g451932; PIDN:AAA16716.1; PID:g451934
R;Neipel, F.; Ellinger, K.; Fleckenstein, B.
J. Virol. 66, 3918-3924, 1992
A;Title: Gene for the major antigenic structural protein (p100) of human
 herpesvirus 6.
A;Reference number: A42533; MUID:92260671; PMID:1374813
A;Accession: A42533
A;Molecule type: DNA
A;Residues: 2-871 <NEI>
A;Cross-references: GB:M87287; NID:g330673; PIDN:AAA46012.1; PID:g330674
C;Genetics:
A;Gene: P1LF1
C;Superfamily: human herpesvirus large structural phosphoprotein; large
 structural phosphoprotein homology
C;Keywords: phosphoprotein
F;7-368/Domain: large structural phosphoprotein homology <CLS>

  XPBE12  Length: 871  October 5, 2004 14:19  Type: P  Check: 1901  ..

     1  NMDLQRHPIP FAWLDRDKVE RLTDFLSNLE RLDNVDLREH PHVTNSCVVR
    51  EGDDVDDLKT LYNLLVLWLM YHYVLSKRKP DYNAIWQDIT KLQSVVNEYL
   101  NSKGLNKGIF ENMFTNKEKF ESQFSDINRA LLRLGNFIKW GSNVAIDTPY
   151  VNLTAEDSSE IENNLQDAEK NMLWYTVYNI NDPWDENGYL ITSINKLIYL
   201  GKLFLALTQS WSKLEKVAMS QIVITQNHLS GHLRRHDNFN IVYSHRVLQT
   251  PLTGQRVESF LKIITSDYDI IKSSLESHSA SKAFSMSEIG PNSLMDFVPL
   301  RGDIHSNLTL PSMSIDTKKS SLDPARLKKS NSRSLDSFLR MQRQPKFLEL
   351  DSVDNAGEKI LLKEATLGGE NVKATTPASS VSLMSGVESP SSFTSTNLDL
   401  PLSSFTSTNL DLRDKSHGNY KIGPSGILDF NVKFPPNAQL NTNGVDLLQD
   451  KTSIGSPSSG ITDVVNGFAN LNLHQNKSNV SPPWSRNTAA NADFLDPVHR
   501  FVPEQTGTPF VLNNSDVAGS EAKHTTYSTE TGVSPRNVFL IKDLRGKDGF
   551  RKQKQSDIPK SLTKERNDKA IMHSREVTGD SGDATETVGA RNSPALRKIK
   601  QANDFFAGLN KKNDRDVLRG GKGNSKDLHS GGNAKKKEMS GKFNDDKEMT
   651  RNGQEPSRSL MGDARNAGDE QYIQAGLGQR VNNLLSQFTN LISLGEKGIE
   701  DILQNQRGTE LKLATENKSG RESEEANVEK ILEVSNPQDM FKNFRLQNDL
   751  DSVQSPFRLP DADLSRELDS ASFKDALDLK LPGNGEREID LALEKVKVGE
   801  TETSDLKVGQ DESFVPAQLM KVETPEEKDD IIEQMVLRIR QDGETDENTV
   851  SGPGVAESLD IEAKGESAIA S
```

```
=> fil reg; d que l3
FILE 'REGISTRY' ENTERED AT 14:06:27 ON 05 OCT 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file
provided by InfoChem.

STRUCTURE FILE UPDATES:    4 OCT 2004  HIGHEST RN 756793-93-8
DICTIONARY FILE UPDATES:   4 OCT 2004  HIGHEST RN 756793-93-8

TSCA INFORMATION NOW CURRENT THROUGH MAY 21, 2004

  Please note that search-term pricing does apply when
  conducting SmartSELECT searches.

Crossover limits have been increased.  See HELP CROSSOVER for details.

Experimental and calculated property data are now available. For more
information enter HELP PROP at an arrow prompt in the file or refer
to the file summary sheet on the web at:
http://www.cas.org/ONLINE/DBSS/registryss.html


L3            39 SEA FILE=REGISTRY ABB=ON  NQTQPE[RQSHYE][GSQITNP][DFTSPLI][NSKM
                 TD][NDKTE][LFRI][TSN]R[IVA]VGGQE/SQSP


=> d rn cn sql kwic nte lc l3 1-39; fil capl uspatf toxcenter; s l3

L3   ANSWER 1 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   701356-61-8  REGISTRY  Use Registry # to match sequence to citation
CN   Soft tissue sarcoma-associated protein (human clone WO2004048938-SEQID-
     261) (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   261: PN: WO2004048938 SEQID: 261 claimed protein
SQL  488
SQL = sequence length
SEQ    201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
                                 ========== ==========
HITS AT:   221-240

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
LC   STN Files:   CA, CAPLUS, TOXCENTER

L3   ANSWER 2 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   674849-48-0  REGISTRY
CN   Protein DITHP (diagnostic and therapeutic protein) (human Incyte clone
     1109930.PT109p) (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   2899: PN: WO2004023973 SEQID: 2899 claimed protein
SQL  406

SEQ    201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
                                 ========== ==========
HITS AT:   221-240
LC   STN Files:   CA, CAPLUS, TOXCENTER

L3   ANSWER 3 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   674849-47-9  REGISTRY
CN   Protein DITHP (diagnostic and therapeutic protein) (human Incyte clone
```

1109930.PT108p) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 2898: PN: WO2004023973 SEQID: 2898 claimed protein
SQL 408

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240
LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 4 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN **674849-46-8** REGISTRY
CN Protein DITHP (diagnostic and therapeutic protein) (human Incyte clone 1109930.PT107p) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 2897: PN: WO2004023973 SEQID: 2897 claimed protein
SQL 427

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240
NTE

type	location	description
uncommon	Aaa-295 - -	

LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 5 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN **674849-45-7** REGISTRY
CN Protein DITHP (diagnostic and therapeutic protein) (human Incyte clone 1109930.PT101p) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 2896: PN: WO2004023973 SEQID: 2896 claimed protein
SQL 437

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240
NTE

type	location	description
uncommon	Aaa-269 - -	

LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 6 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN **674849-44-6** REGISTRY
CN Protein DITHP (diagnostic and therapeutic protein) (human Incyte clone 1109930.PT100p) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 2895: PN: WO2004023973 SEQID: 2895 claimed protein
SQL 437

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240
LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 7 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN

RN 631960-24-2 REGISTRY
CN 290: PN: WO03099862 PAGE: 70 unclaimed sequence (9CI) (CA INDEX NAME)
SQL 449

SEQ 151 SSGEAPDSIT WKPYDAADLD PTENPFDLLD FNQTQPERGD NNLTRIVGGQ
========== ==========
201 ECKDGECPWQ ALLINEENEG FCGGTILSEF YILTAAHCLY QAKRFKVRVG
=
HITS AT: 182-201
LC STN Files: CA, CAPLUS, TOXCENTER, USPATFULL

L3 ANSWER 8 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 623046-41-3 REGISTRY
CN GenBank AAH46125 (9CI) (CA INDEX NAME)
OTHER NAMES:
CN GenBank AAH46125 (TRANSLATED FROM: GenBank BC046125)
SQL 488

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240

RELATED SEQUENCES AVAILABLE WITH SEQLINK

L3 ANSWER 9 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 612559-84-9 REGISTRY
CN 8: PN: WO03082914 FIGURE: 22 unclaimed sequence (9CI) (CA INDEX NAME)
SQL 488

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240
LC STN Files: CA, CAPLUS, TOXCENTER, USPATFULL

L3 ANSWER 10 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 569383-76-2 REGISTRY
CN Protein PMMM-30 (protein modification and maintenance molecule) (human
Incyte clone 7509223) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 30: PN: WO03063688 SEQID: 30 claimed protein
SQL 442

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240
LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 11 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 569383-75-1 REGISTRY
CN Protein PMMM-29 (protein modification and maintenance molecule) (human
Incyte clone 7509140) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 29: PN: WO03063688 SEQID: 29 claimed protein
SQL 377

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240
LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 12 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 569383-74-0 REGISTRY

CN Protein PMMM-28 (protein modification and maintenance molecule) (human Incyte clone 7509113) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 28: PN: WO03063688 SEQID: 28 claimed protein
SQL 444

SEQ 151 PDSITWKPYD AADLDPTENP FDLLDFNQTQ PERGDNNLTR IVGGQECKDG
==== ========== ======
HITS AT: 177-196
LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 13 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN **487495-44-3** REGISTRY
CN GenBank CAB69368 (9CI) (CA INDEX NAME)
OTHER NAMES: GenBank record printed at end of search, beginning at yellow tab
CN GenBank CAB69368 (Translated from: GenBank A86886)
SQL 488

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240

RELATED SEQUENCES AVAILABLE WITH SEQLINK

L3 ANSWER 14 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 487494-83-7 REGISTRY
CN GenBank CAB69367 (9CI) (CA INDEX NAME)
OTHER NAMES:
CN GenBank CAB69367 (Translated from: GenBank A86859)
SQL 488

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240

RELATED SEQUENCES AVAILABLE WITH SEQLINK

L3 ANSWER 15 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 481289-90-1 REGISTRY
CN GenBank AAA52764 (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 2162: PN: WO03091391 FIGURE: 20 unclaimed protein
CN 386: PN: WO03091391 FIGURE: 17 unclaimed protein
CN GenBank AAA52764 (Translated from: GenBank L29433)
SQL 488

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240

RELATED SEQUENCES AVAILABLE WITH SEQLINK
LC STN Files: CA, CAPLUS, USPATFULL

L3 ANSWER 16 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 481286-57-1 REGISTRY
CN GenBank AAA52490 (9CI) (CA INDEX NAME)
OTHER NAMES:
CN GenBank AAA52490 (Translated from: GenBank K03194)
SQL 476

SEQ 201 NPFDLLDFNQ TQPERGDNNL TRIVGGQECK DGECPWQALL INEENEGFCG
== ========== ========

HITS AT: 209-228

L3 ANSWER 17 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 481286-53-7 REGISTRY
CN Protein (human gene F10) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 2023: PN: US20040009481 TABLE: 1 claimed protein
CN GenBank AAA52486
CN GenBank AAA52486 (Translated from: GenBank K01886)
SQL 371

SEQ 101 FDLLDFNQTQ PERGDNNLTR IVGGQECKDG ECPWQALLIN EENEGFCGGT
==== ========== ======
HITS AT: 107-126
LC STN Files: CA, CAPLUS, TOXCENTER, USPATFULL

L3 ANSWER 18 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 481286-00-4 REGISTRY
CN GenBank AAA52421 (9CI) (CA INDEX NAME)
OTHER NAMES:
CN GenBank AAA52421 (Translated from: GenBank M57285)
SQL 488

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240

RELATED SEQUENCES AVAILABLE WITH SEQLINK

L3 ANSWER 19 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 481281-54-3 REGISTRY
CN GenBank AAA51984 (9CI) (CA INDEX NAME)
OTHER NAMES:
CN GenBank AAA51984 (Translated from: GenBank M22613)
SQL 467

SEQ 201 DFNQTQPERG DNNLTRIVGG QECKDGECPW QALLINEENE GFCGGTILSE
======== ========== ==
HITS AT: 203-222

L3 ANSWER 20 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 479878-50-7 REGISTRY
CN GenBank AAM19347 (9CI) (CA INDEX NAME)
OTHER NAMES:
CN GenBank AAM19347 (Translated from: GenBank AF503510)
SQL 488

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
========== ==========
HITS AT: 221-240

RELATED SEQUENCES AVAILABLE WITH SEQLINK
LC STN Files: CA, CAPLUS

L3 ANSWER 21 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN **325762-37-6** REGISTRY
CN Blood-coagulation factor X, prepro-[227-serine,228-glutamine,229-threonine,230-serine,231-lysine,232-leucine,233-threonine] (human liver) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 8: PN: WO0110896 SEQID: 2 claimed protein
SQL 488

```
SEQ    201 KPYDAADLDP TENPFDLLDF NQTQPESQTS KLTRIVGGQE CKDGECPWQA
                                  ========== ==========
HITS AT:   221-240
LC   STN Files:   CA, CAPLUS, TOXCENTER

L3   ANSWER 22 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   325762-36-5  REGISTRY
CN   Blood-coagulation factor X, prepro-[227-glutamine,228-serine,229-
     phenylalanine,230-asparagine,231-aspartic acid,232-phenylalanine,233-
     threonine] (human liver) (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   7: PN: WO0110896 SEQID: 2 claimed protein
SQL  488

SEQ    201 KPYDAADLDP TENPFDLLDF NQTQPEQSFN DFTRIVGGQE CKDGECPWQA
                                  ========== ==========
HITS AT:   221-240
LC   STN Files:   CA, CAPLUS, TOXCENTER

L3   ANSWER 23 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   325762-30-9  REGISTRY
CN   Blood-coagulation factor X, prepro-(human liver) (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   2: PN: WO0110896 SEQID: 2 claimed protein
SQL  488

SEQ    201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
                                  ========== ==========
HITS AT:   221-240

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
LC   STN Files:   CA, CAPLUS, TOXCENTER

L3   ANSWER 24 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   212622-93-0  REGISTRY
CN   Blood-coagulation factor X, pro- [347-asparagine] (human) (9CI)  (CA INDEX
     NAME)
SQL  448

SEQ    151 SGEAPDSITW KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE
                                             ========== ==========
HITS AT:   181-200
LC   STN Files:   CA, CAPLUS

L3   ANSWER 25 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   212380-22-8  REGISTRY
CN   Blood-coagulation factor X, prepro- [469-lysine] (human fibroblast) (9CI)
     (CA INDEX NAME)
SQL  488

SEQ    201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
                                  ========== ==========
HITS AT:   221-240
LC   STN Files:   CA, CAPLUS, USPATFULL

L3   ANSWER 26 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   212380-18-2  REGISTRY
CN   1-469-Blood-coagulation factor X, prepro- (human fibroblast) (9CI)  (CA
     INDEX NAME)
SQL  469
```

```
SEQ    201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
                                  ========== ==========
HITS AT:   221-240
LC   STN Files:   CA, CAPLUS, USPATFULL

L3   ANSWER 27 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   212379-48-1  REGISTRY
CN   Blood-coagulation factor X, prepro- [231-aspartic acid,232-
     phenylalanine,235-valine] (human fibroblast) (9CI)  (CA INDEX NAME)
SQL  488

SEQ    201 KPYDAADLDP TENPFDLLDF NQTQPERGDN DFTRVVGGQE CKDGECPWQA
                                  ========== ==========
HITS AT:   221-240
LC   STN Files:   CA, CAPLUS, USPATFULL

L3   ANSWER 28 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   174294-63-4  REGISTRY
CN   Blood-coagulation factor X, pro- [282-asparagine,379-alanine] (human
     reduced) (9CI)  (CA INDEX NAME)
SQL  448

SEQ    151 SGEAPDSITW KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE
                                             ========== ==========
HITS AT:   181-200
LC   STN Files:   CA, CAPLUS

L3   ANSWER 29 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   174294-62-3  REGISTRY
CN   Blood-coagulation factor X, pro- [282-asparagine] (human) (9CI)  (CA INDEX
     NAME)
SQL  448

SEQ    151 SGEAPDSITW KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE
                                             ========== ==========
HITS AT:   181-200
LC   STN Files:   CA, CAPLUS

L3   ANSWER 30 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   174294-61-2  REGISTRY
CN   Blood-coagulation factor X, pro- [379-alanine] (human clone pBNX) (9CI)
     (CA INDEX NAME)
SQL  448

SEQ    151 SGEAPDSITW KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE
                                             ========== ==========
HITS AT:   181-200
LC   STN Files:   CA, CAPLUS

L3   ANSWER 31 OF 39  REGISTRY  COPYRIGHT 2004 ACS on STN
RN   138824-08-5  REGISTRY
CN   L-Cysteine, L-prolyl-L-phenylalanyl-L-.alpha.-aspartyl-L-leucyl-L-leucyl-L-
     .alpha.-aspartyl-L-phenylalanyl-L-asparaginyl-L-glutaminyl-L-threonyl-L-
     glutaminyl-L-prolyl-L-.alpha.-glutamyl-L-arginylglycyl-L-.alpha.-aspartyl-
     L-asparaginyl-L-asparaginyl-L-leucyl-L-threonyl-L-arginyl-L-isoleucyl-L-
     valylglycylglycyl-L-glutaminyl-L-.alpha.-glutamyl- (9CI)  (CA INDEX NAME)
SQL  28

SEQ      1 PFDLLDFNQT QPERGDNNLT RIVGGQEC
               === ========== =======
HITS AT:   8-27
LC   STN Files:   CA, CAPLUS
```

L3 ANSWER 32 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 103850-95-9 REGISTRY
CN Blood-coagulation factor X, pro- (human fibroblast protein moiety reduced) (9CI) (CA INDEX NAME)
SQL 448

SEQ 151 SGEAPDSITW KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE
==========

HITS AT: 181-200
LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 33 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 103850-94-8 REGISTRY
CN Blood-coagulation factor X, pro- (human fibroblast protein moiety) (9CI) (CA INDEX NAME)
SQL 445,306,139

SEQ 1 SVAQATSSSG EAPDSITWKP YDAADLDPTE NPFDLLDFNQ TQPERGDNNL
==

51 TRIVGGQECK DGECPWQALL INEENEGFCG GTILSEFYIL TAAHCLYQAK
========

HITS AT: 39-58

NTE multichain
modified (modifications unspecified)

type	------ location ------		description
bridge	Cys-59	- Cys-64	disulfide bridge
bridge	Cys-79	- Cys-95	disulfide bridge
bridge	Cys-160	- Cys-132'	disulfide bridge
bridge	Cys-208	- Cys-222	disulfide bridge
bridge	Cys-233	- Cys-261	disulfide bridge
bridge	Cys-50'	- Cys-61'	disulfide bridge
bridge	Cys-55'	- Cys-70'	disulfide bridge
bridge	Cys-72'	- Cys-81'	disulfide bridge
bridge	Cys-89'	- Cys-100'	disulfide bridge
bridge	Cys-96'	- Cys-109'	disulfide bridge
bridge	Cys-111'	- Cys-124'	disulfide bridge
uncommon	Gla-6'	-	-
uncommon	Gla-7'	-	-
uncommon	Gla-14'	-	-
uncommon	Gla-16'	-	-
uncommon	Gla-19'	-	-
uncommon	Gla-20'	-	-
uncommon	Gla-25'	-	-
uncommon	Gla-26'	-	-
uncommon	Gla-29'	-	-
uncommon	Gla-32'	-	-
uncommon	Gla-39'	-	-

LC STN Files: CA, CAPLUS

L3 ANSWER 34 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 103850-93-7 REGISTRY
CN Blood-coagulation factor X, prepro- (human fibroblast protein moiety reduced) (9CI) (CA INDEX NAME)
SQL 488

SEQ 201 KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE CKDGECPWQA
==========

HITS AT: 221-240

RELATED SEQUENCES AVAILABLE WITH SEQLINK
LC STN Files: CA, CAPLUS, TOXCENTER, USPATFULL

L3 ANSWER 35 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN **103850-92-6** REGISTRY
CN Blood-coagulation factor X, prepro- (human fibroblast protein moiety) (9CI) (CA INDEX NAME)
SQL 485,306,179

SEQ 1 SVAQATSSSG EAPDSITWKP YDAADLDPTE NPFDLLDFNQ TQPERGDNNL
== ==========
51 TRIVGGQECK DGECPWQALL INEENEGFCG GTILSEFYIL TAAHCLYQAK
========
HITS AT: 39-58

NTE multichain
modified (modifications unspecified)

type	------ location ------		description
bridge	Cys-59	- Cys-64	disulfide bridge
bridge	Cys-79	- Cys-95	disulfide bridge
bridge	Cys-160	- Cys-172'	disulfide bridge
bridge	Cys-208	- Cys-222	disulfide bridge
bridge	Cys-233	- Cys-261	disulfide bridge
bridge	Cys-90'	- Cys-101'	disulfide bridge
bridge	Cys-95'	- Cys-110'	disulfide bridge
bridge	Cys-112'	- Cys-121'	disulfide bridge
bridge	Cys-129'	- Cys-140'	disulfide bridge
bridge	Cys-136'	- Cys-149'	disulfide bridge
bridge	Cys-151'	- Cys-164'	disulfide bridge
uncommon	Gla-46'	-	-
uncommon	Gla-47'	-	-
uncommon	Gla-54'	-	-
uncommon	Gla-56'	-	-
uncommon	Gla-59'	-	-
uncommon	Gla-60'	-	-
uncommon	Gla-65'	-	-
uncommon	Gla-66'	-	-
uncommon	Gla-69'	-	-
uncommon	Gla-72'	-	-
uncommon	Gla-79'	-	-

LC STN Files: CA, CAPLUS

L3 ANSWER 36 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 103850-91-5 REGISTRY
CN Blood-coagulation factor X (human fibroblast heavy chain protein moiety reduced) (9CI) (CA INDEX NAME)
SQL 306

SEQ 1 SVAQATSSSG EAPDSITWKP YDAADLDPTE NPFDLLDFNQ TQPERGDNNL
== ==========
51 TRIVGGQECK DGECPWQALL INEENEGFCG GTILSEFYIL TAAHCLYQAK
========
HITS AT: 39-58
LC STN Files: CA, CAPLUS, TOXCENTER

L3 ANSWER 37 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN **103813-13-4** REGISTRY

CN Blood-coagulation factor X (human clone 4C heavy chain protein moiety reduced) (9CI) (CA INDEX NAME)
SQL 303

SEQ 1 SVAQATSSSG EAPDSITWKP YDAADLDPTE NPFDLLDFNQ TQPERGDNNL
== ==========
51 TRIVGGQECK DGECPWQALL INEENEGFCG GTILSEFYIL TAAHCLYQAK
========
HITS AT: 39-58
LC STN Files: CA, CAPLUS

L3 ANSWER 38 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 91930-66-4 REGISTRY
CN Blood-coagulation factor X, pro- (human protein moiety reduced) (9CI) (CA INDEX NAME)
SQL 448

SEQ 151 SGEAPDSITW KPYDAADLDP TENPFDLLDF NQTQPERGDN NLTRIVGGQE
========== ==========
HITS AT: 181-200
LC STN Files: CA, CAPLUS

L3 ANSWER 39 OF 39 REGISTRY COPYRIGHT 2004 ACS on STN
RN 91930-64-2 REGISTRY
CN Blood-coagulation factor X (human heavy chain reduced) (9CI) (CA INDEX NAME)
SQL 306

SEQ 1 SVAQATSSSG EAPDSITWKP YDAADLDPTE NPFDLLDFNQ TQPERGDNNL
== ==========
51 TRIVGGQECK DGECPWQALL INEENEGFCG GTILSEFYIL TAAHCLYQAK
========
HITS AT: 39-58
LC STN Files: CA, CAPLUS

FILE 'CAPLUS' ENTERED AT 14:07:05 ON 05 OCT 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'USPATFULL' ENTERED AT 14:07:05 ON 05 OCT 2004
CA INDEXING COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'TOXCENTER' ENTERED AT 14:07:05 ON 05 OCT 2004
COPYRIGHT (C) 2004 ACS

L6 36 L3 Registry file answer set crossed into bibliographic files to get citations

=> dup rem l6
PROCESSING COMPLETED FOR L6
L7 26 DUP REM L6 (10 DUPLICATES REMOVED)
ANSWERS '1-19' FROM FILE CAPLUS
ANSWERS '20-26' FROM FILE USPATFULL

=> d ibib ed ab hitrn 1-26; fil hom

L7 ANSWER 1 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 1
ACCESSION NUMBER: 2004:471053 CAPLUS
DOCUMENT NUMBER: 141:37227

TITLE: Gene expression profiles for detecting soft tissue sarcomas and compositions and methods of screening for soft tissue sarcoma modulators
INVENTOR(S): Aziz, Natasha; Ginsburg, Wendy M.; Zlotnik, Albert
PATENT ASSIGNEE(S): Protein Design Labs, Inc., USA
SOURCE: PCT Int. Appl., 210 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 4
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2004048938	A2	20040610	WO 2003-US38193	20031126

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ
RW: BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: US 2002-429739P P 20021126
ED Entered STN: 10 Jun 2004
AB Described herein are methods and compns. that can be used for diagnosis and treatment of soft tissue sarcoma cancer phenotypes and soft tissue sarcoma cancer-assocd. diseases. Also described herein are methods that can be used to identify modulators of soft tissue sarcoma cancer. The Eos/Affymetrix Hu03 Genechip microarray was used to identify up-regulated genes in various human soft tissue sarcomas: 523 genes up-regulated in chondrosarcoma, 763 genes in dermatofibrosarcoma, 625 genes in fibrosarcoma, 906 genes in liposarcoma, 595 genes in synovial sarcoma, 977 genes in rhabdomyosarcoma, and 1078 genes in malignant fibrous histiocytoma. [This abstr. record is one of four records for this document necessitated by the large no. of index entries required to fully index the document and publication system constraints.].
IT **701356-61-8** - Use Registry # to match citation to sequence
RL: BSU (Biological study, unclassified); DGN (Diagnostic use); PRP (Properties); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(amino acid sequence; gene expression profiles for detecting soft tissue sarcomas and compns. and methods of screening for soft tissue sarcoma modulators)

L7 ANSWER 2 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 2
ACCESSION NUMBER: 2004:250754 CAPLUS
DOCUMENT NUMBER: 140:282451
TITLE: Human protein and cDNA sequences for diagnostics and therapeutics
INVENTOR(S): Schmidt, Jeanette P.; Wright, Rachel J.; Bruns, Christopher M.; Marjanovic, Mirjana M.; Shen, Fan; Harthshorne, Toinette A.; Suchorolski, Martin T.; Altus, Christina M.; Pitts, Steven J.; Elder, Linda V.; Mooney, Elizabeth M.; Delegeane, Angelo M.; Panesar, Iqbal S.; Banville, Steven C.; Reddy, Thirupathi P.; Stevens, Kristian A.; Blanchard, John L.; Panzer, Scott R.; Wang, Xinhao; Au, Alan P.; Gerstin, Edward H., Jr.; Peralta, Careyna H.; Anderson, Scott B.; Rioux, Pierre; Shen, Edward J.;

Wu, Mingham C.; Stuve, Laura L.; Lagace, Robert E.; Spiro, Peter A.; Stewart, Elizabeth A.; Wingrove, James; Vitt, Ursula A.; Kirton, Edward S.; Xu, Yuming; Kwong, Mary; Policky, Jennifer L.; Hurwitz, Bonnie L.; Ma, Yan; Jackson, Jennifer L.; Gietzen, Darryl; Patury, Srikanth; Shi, Xiaobing; Suarez, Charlyn J.

PATENT ASSIGNEE(S): Incyte Corporation, USA
SOURCE: PCT Int. Appl., 190 pp.
CODEN: PIXXD2

DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2004023973	A2	20040325	WO 2003-US28227	20030912

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: US 2002-410259P P 20020912
US 2002-410260P P 20020912

ED Entered STN: 26 Mar 2004

AB The present invention provides 2722 human cDNA sequences for diagnostics and therapeutics (dithp) and the polypeptides (DITHP) encoded by dithp. The invention also provides for the use of dithp, or complements, oligonucleotides, or fragments thereof in diagnostic assays. The invention further provides for vectors and host cells contg. dithp for the expression of DITHP. The invention addnl. provides for the use of isolated and purified DITHP to induce antibodies and to screen libraries of compds. and the use of anti-DITHP antibodies in diagnostic assays. Also provided are microarrays contg. dithp and methods of use.

IT **674849-44-6 674849-45-7 674849-46-8**
674849-47-9 674849-48-0
RL: ANT (Analyte); BSU (Biological study, unclassified); DGN (Diagnostic use); PRP (Properties); THU (Therapeutic use); ANST (Analytical study); BIOL (Biological study); USES (Uses)
(amino acid sequence; human protein and cDNA sequences for diagnostics and therapeutics)

L7 ANSWER 3 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 3
ACCESSION NUMBER: 2004:39697 CAPLUS
DOCUMENT NUMBER: 140:123703
TITLE: Human prostate cancer marker genes associated with various metastatic stages identified by gene profiling, and related compositions, kits, and methods for diagnosis, prognosis and therapy
INVENTOR(S): Schlegel, Robert; Endege, Wilson O.
PATENT ASSIGNEE(S): Millennium Pharmaceuticals, Inc., USA
SOURCE: U.S. Pat. Appl. Publ., 131 pp.
CODEN: USXXCO

DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 5
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.		DATE
US 2004009481	A1	20040115	US 2002-166883		20020611
US 2004009481	A1	20040115	US 2002-166883		20020611
US 2004009481	A1	20040115	US 2002-166883		20020611
US 2004009481	A1	20040115	US 2002-166883		20020611
US 2004009481	A1	20040115	US 2002-166883		20020611
PRIORITY APPLN. INFO.:			US 2001-297285P	P	20010611
			US 2002-166883	A	20020611

ED Entered STN: 16 Jan 2004

AB The invention relates to compns., kits, and methods for diagnosing, staging, prognosing, monitoring and treating human prostate cancers. A variety of marker genes are provided, wherein changes in the levels of expression of one or more of the marker genes is correlated with the presence of prostate cancer. In particular, three sets of the marker genes set, corresponding to 11617 GenBank Accession Nos. (only 2168 new submissions) and 15 SEQ IDs, are identified by transcription profiling using RNA derived from clin. samples, that were expressed at least 2-fold or greater than the normal controls. Using TNM staging approach, these markers are divided to three groups, ones can be used to det. whether prostate cancer has metastasized, or is likely to metastasize, to the liver (M stage); ones can be used to det. whether prostate cancer has metastasized, or is likely to metastasize, to the bone (M stage); and ones can be used to det. whether prostate cancer has metastasized, or is likely to metastasize, to the lymph nodes (N stage and/or M stage). The invention also relates to a kit for assessing the specific type of metastatic prostate cancer, e.g., cancer that has metastasized to the liver, bone or lymph nodes. [This abstr. record is one of three records for this document necessitated by the large no. of index entries required to fully index the document and publication system constraints.].

IT **481286-53-7**, Protein (human gene F10)
RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(amino acid sequence; human prostate cancer marker genes assocd. with various metastatic stages identified by gene profiling, and related compns., kits, and methods for diagnosis, prognosis and therapy)

L7 ANSWER 4 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 4
ACCESSION NUMBER: 2003:951061 CAPLUS
DOCUMENT NUMBER: 140:26964
TITLE: Use of the lantibiotic transport system to secrete foreign proteins into culture medium for purification
INVENTOR(S): Moll, Gert Nikolaas; Leenhouts, Cornelis Johannes; Kuipers, Oscar Paul; Driessen, Arnold Jacob Mathieu
PATENT ASSIGNEE(S): Applied Nanosystems B.V., Neth.
SOURCE: PCT Int. Appl., 109 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003099862	A1	20031204	WO 2003-NL389	20030526
WO 2003099862	C1	20040311		

W: AE, AG, AL, AM, AT, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, CZ, DE, DE, DK, DK, DM, DZ, EC, EE, EE, ES, FI, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA,

ZM, ZW, AM, AZ
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

US 2004009550	A1	20040115	US 2003-360101	20030207
PRIORITY APPLN. INFO.:			EP 2002-77060 A	20020524
			US 2003-360101 A	20030207

ED Entered STN: 07 Dec 2003

AB Methods of using the mechanisms involved in the secretion of lantibiotics to secrete foreign proteins from lantibiotic-producing hosts is described. The method can also be used to secrete lantibiotics before they have undergone post-translational modification, such as dehydration of a serine or a threonine, and/or thioether bridge formation, or to increase the efficiency of secretion of fully processed lantibiotics. A Lactococcus lactis strain lacking the entire nisin A biosynthetic gene cluster was transformed with a plasmid carrying the nisin A structural gene nisA and the transport protein nisT. This transgenic strain efficiently secreted the unmodified nisin A protein, indicating that lanT was sufficient to export the protein. Use of the signal peptide to direct secretion of an angiotensin variant is demonstrated. Use of the transport protein, the lantibiotic signal peptide, and the lantibiotic-modifying dehydrases and cyclases to manuf. novel variants of peptide hormones with modified amino is also demonstrated.

IT **631960-24-2**
RL: PRP (Properties)
(unclaimed sequence; use of the lantibiotic transport system to secrete foreign proteins into culture medium for purifn.)

REFERENCE COUNT: 11 THERE ARE 11 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 5 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 5

ACCESSION NUMBER:	2003:942767 CAPLUS
DOCUMENT NUMBER:	140:40262
TITLE:	Genes expressed in atherosclerotic tissue and their use in diagnosis and pharmacogenetics
INVENTOR(S):	Nevins, Joseph; West, Mike; Goldschmidt, Pascal
PATENT ASSIGNEE(S):	Duke University, USA
SOURCE:	PCT Int. Appl., 408 pp. CODEN: PIXXD2
DOCUMENT TYPE:	Patent
LANGUAGE:	English
FAMILY ACC. NUM. COUNT:	3
PATENT INFORMATION:	

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003091391	A2	20031106	WO 2002-XB38221	20021112

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

WO 2003091391	A2	20031106	WO 2002-US38221	20021112

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN,

MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.:

US 2002-374547P	P	20020423
US 2002-420784P	P	20021024
US 2002-421043P	P	20021025
US 2002-424680P	P	20021108
WO 2002-US38221	A	20021112

ED Entered STN: 04 Dec 2003

AB Genes whose expression is correlated with an determinant of an atherosclerotic phenotype are provided. Also provided are methods of using the subject atherosclerotic determinant genes in diagnosis and treatment methods, as well as drug screening methods. In addn., reagents and kits thereof that find use in practicing the subject methods are provided. Also provided are methods of detg. whether a gene is correlated with a disease phenotype, where correlation is detd. using a Bayesian anal. [This abstr. record is one of three records for this document necessitated by the large no. of index entries required to fully index the document and publication system constraints.].

IT **481289-90-1**, GenBank AAA52764
RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(amino acid sequence; genes expressed in atherosclerotic tissue and their use in diagnosis and pharmacogenetics)

L7 ANSWER 6 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 6

ACCESSION NUMBER: 2003:796743 CAPLUS
DOCUMENT NUMBER: 139:318435
TITLE: Prothrombin activating proteins SVPs (snake venom proteases) from Australian snakes, nucleic acid sequences encoding same, and methods to promote hemostasis and prevent blood loss
INVENTOR(S): Masci, Paul Pantaleone; De Jersey, John; Lavin, Martin
PATENT ASSIGNEE(S): The University of Queensland, Australia
SOURCE: PCT Int. Appl., 173 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003082914	A1	20031009	WO 2003-AU406	20030403

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

US 2004043017	A1	20040304	US 2003-406031	20030402

PRIORITY APPLN. INFO.:

AU 2002-1483	A	20020403
AU 2003-901033	A	20030307

ED Entered STN: 10 Oct 2003

AB The invention relates to prothrombin activating proteins, referred to as "snake venom proteases" or SVPs, from venom of Australian snakes, and nucleic acid sequences encoding same. The snake venom proteases share certain amino acid sequences similarity to that of the human factor Xa and trocarin. However, the snake venom proteases of the invention are complete or partially complete prothrombin activators and they can process prothrombin to thrombin in the absence of cofactors such as calcium, phospholipids and/or factor Va. Examples of complete SVP's include SVP's from Brown, coastal Taipan, or inland Taipan snakes, and of the partially complete SVP's include SVP's from Red Belly Black, Tiger, or Rough Scale snakes. These SVP's appear to include an internal domain, which makes them independent of host supplied factor Va. In addn., preferred SVP's of the invention can cleave descarboxy prothrombin, which is a poor substrate for human factor X. This invention also relates to methods of making and using the SVPs, e.g., to promote hemostasis and prevent blood loss such as during surgery or for treatment of wounds resulting from accidents and other types of injury or trauma.

IT **612559-84-9**
RL: PRP (Properties)
(unclaimed sequence; prothrombin activating proteins SVPs (snake venom proteases) from Australian snakes, nucleic acid sequences encoding same, and methods to promote hemostasis and prevent blood loss)

REFERENCE COUNT: 3 THERE ARE 3 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 7 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 7

ACCESSION NUMBER: 2003:610168 CAPLUS
DOCUMENT NUMBER: 139:145036
TITLE: Novel human protein modification and maintenance molecules and their cDNAs and therapeutic use therefor
INVENTOR(S): Hafalia, April J. A.; Li, Joana X.; Gorvad, Ann E.; Chawla, Narinder K.; Sprague, William W.; Lee, Soo Yeun; Chang, Hsin-Ru; Elliott, Vicki S.; Ramkumar, Jayalaxmi; Khare, Reena; Emerling, Brooke M.; Kable, Amy E.; Tang, Y. Tom; Yue, Henry; Gietzen, Kimberly J.; Lee, Sally; Swarnakar, Anita; Baughn, Mariah R.; Wilson, Amy D.; Jin, Pei; Chien, David; Hawkins, Phillip R.; Jiang, Xin; Jackson, Alan A.; Bhatia, Umesh; Burrill, John D.; Blake, Julie J.; Ho, Anne; Zheng, Wenjin; Ison, Craig H.; Marquis, Joseph P.; Tran, Uyen K.; Lal, Preeti G.; Warren, Bridget A.; Xu, Yuming; Honchell, Cynthia D.; Becha, Shanya D.; Lehr-Mason, Patricia M.
PATENT ASSIGNEE(S): Incyte Genomics, Inc., USA
SOURCE: PCT Int. Appl., 405 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 71
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003063688	A2	20030807	WO 2003-US2500	20030123

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, CH,

CY, DE, DK, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: US 2002-351928P P 20020125
US 2002-359903P P 20020225
US 2002-366837P P 20020321

ED Entered STN: 08 Aug 2003

AB The invention provides 58 human protein modification and maintenance mol. (designated PMMM-x, x from 1-58) and polynucleotides which identify and encode PMMM. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders assocd. with aberrant expression of PMMM.

IT **569383-74-0P 569383-75-1P 569383-76-2P**
RL: ANT (Analyte); BPN (Biosynthetic preparation); DGN (Diagnostic use); PRP (Properties); THU (Therapeutic use); ANST (Analytical study); BIOL (Biological study); PREP (Preparation); USES (Uses)
(amino acid sequence; novel human protein modification and maintenance mols. and their cDNAs and therapeutic use therefor)

L7 ANSWER 8 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 8
ACCESSION NUMBER: 2001:115178 CAPLUS
DOCUMENT NUMBER: 134:168320
TITLE: Factor X substitution mutant with an improved ability to be activated
INVENTOR(S): Himmelspach, Michele; Schlokat, Uwe
PATENT ASSIGNEE(S): Baxter A.-G., Austria
SOURCE: PCT Int. Appl., 50 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001010896	A2	20010215	WO 2000-EP7631	20000807
WO 2001010896	A3	20020711		
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
AT 9901377	A	20020715	AT 1999-1377	19990810
AT 410216	B	20030325		
EP 1238065	A2	20020911	EP 2000-949465	20000807
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL				

PRIORITY APPLN. INFO.: AT 1999-1377 A 19990810
WO 2000-EP7631 W 20000807

OTHER SOURCE(S): MARPAT 134:168320

ED Entered STN: 15 Feb 2001

AB The invention relates to factor Xa analogs with a modified protease cleavage site, comprising a substitution of a min. of one of the amino acid between Glu226 and Arg234 and possibly Ile235 in the region of activation peptide. These modified cleavage sites in the region of activation peptide change protease specificity and facilitate factor XIa cleavage of the precursor. The invention also relates to prepns. contg.

said factor Xa analogs and methods for the prodn. thereof. The prepro-factor X analogs may be used to produce high-purity factor X for use as coagulants.

IT **325762-30-9P 325762-36-5P 325762-37-6P**
RL: BPN (Biosynthetic preparation); BPR (Biological process); BSU (Biological study, unclassified); PUR (Purification or recovery); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); PROC (Process); USES (Uses)
(amino acid sequence; factor X substitution mutant with improved ability to be activated)

L7 ANSWER 9 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 9
ACCESSION NUMBER: 1991:649467 CAPLUS
DOCUMENT NUMBER: 115:249467
TITLE: Cloning and expression in COS-1 cells of a full-length cDNA encoding human coagulation factor X
AUTHOR(S): Messier, Terri L.; Pittman, Debra D.; Long, George L.; Kaufman, Randal J.; Church, William R.
CORPORATE SOURCE: Dep. Biochem., Univ. Vermont, Burlington, VT, 05405, USA
SOURCE: Gene (1991), 99(2), 291-4
CODEN: GENED6; ISSN: 0378-1119
DOCUMENT TYPE: Journal
LANGUAGE: English
ED Entered STN: 14 Dec 1991
AB A 1.5-kb cDNA (FX) encoding full-length human coagulation factor X was isolated from a human fetal liver cDNA library. The identity of the insertion in a selected phage .lambda. clone was confirmed to be FX by nucleotide (nt) sequence anal. and restriction mapping. This FX cDNA clone contained 1467 bp of coding sequence, no 5'-untranslated sequence, short 3'-untranslated sequence of 10 nt and a poly(A) tail at the 3'-end. The FX cDNA was inserted into a mammalian expression vector and transfected into COS-1 monkey kidney cells. Media from transfected cells showed evidence of factor X antigen and, following addn. of Russel's viper venom factor X activator, enhanced amidolytic activity toward a synthetic peptide p-nitroanilide substrate. Immunopptn. with an anti-factor X monoclonal antibody of [35S]methionine-labeled cell-conditioned media showed evidence of polypeptides of 74, 55, and 17 kDa, as detd. by SDS-PAGE followed by autoradiog. Together, these results indicate that an active factor X can be successfully expressed in a recombinant DNA expression system. This approach will allow the systematic structure/function investigation of this important blood clotting enzyme.
IT **103850-91-5**, Blood-coagulation factor X (human fibroblast heavy chain protein moiety reduced) **103850-93-7 103850-95-9**
RL: PRP (Properties)
(amino acid sequence of)

L7 ANSWER 10 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 2004:524791 CAPLUS
DOCUMENT NUMBER: 141:237449
TITLE: Evidence for substantial fine-scale variation in recombination rates across the human genome
AUTHOR(S): Crawford, Dana C.; Bhangale, Tushar; Li, Na; Hellenthal, Garrett; Rieder, Mark J.; Nickerson, Deborah A.; Stephens, Matthew
CORPORATE SOURCE: Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA
SOURCE: Nature Genetics (2004), 36(7), 700-706
CODEN: NGENEC; ISSN: 1061-4036
PUBLISHER: Nature Publishing Group
DOCUMENT TYPE: Journal
LANGUAGE: English

ED Entered STN: 30 Jun 2004
AB Characterizing fine-scale variation in human recombination rates is important, both to deepen understanding of the recombination process and to aid the design of disease assocn. studies. Current genetic maps show that rates vary on a megabase scale, but studying finer-scale variation using pedigrees is difficult. Sperm-typing expts. have characterized regions where crossovers cluster into 1-2-kb hot spots, but tech. difficulties limit the no. of studies. An alternative is to use population variation to infer fine-scale characteristics of the recombination process. Several surveys reported 'block-like' patterns of diversity, which may reflect fine-scale recombination rate variation, but limitations of available methods made this impossible to assess. Here, we applied a new statistical method, which overcomes these limitations, to infer patterns of fine-scale recombination rate variation in 74 genes. We found extensive rate variation both within and among genes. In particular, recombination hot spots are a common feature of the human genome; 47% (35 of 74) of genes showed substantive evidence for a hot spot, and many more showed evidence for some rate variation. No primary sequence characteristics are consistently assocd. with precise hot-spot location, although G+C content and nucleotide diversity are correlated with local recombination rate.
IT **479878-50-7**, GenBank AAM19347
RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(amino acid sequence; evidence for substantial fine-scale variation in recombination rates across the human genome)
REFERENCE COUNT: 56 THERE ARE 56 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 11 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1998:612199 CAPLUS
DOCUMENT NUMBER: 129:227497
TITLE: Factor X substitution mutant with reduced affinity for factor Va
INVENTOR(S): Miletich, Joseph P.
PATENT ASSIGNEE(S): Washington University, USA
SOURCE: PCT Int. Appl., 42 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.		DATE
WO 9839456	A1	19980911	WO 1998-US3939		19980305
W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM					
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG					
AU 9866738	A1	19980922	AU 1998-66738		19980305
PRIORITY APPLN. INFO.:			US 1997-40047P	P	19970307
			WO 1998-US3939	W	19980305

ED Entered STN: 28 Sep 1998
AB A Factor X variant is disclosed which has a single substitution in which the arginine residue at amino acid position 347 of human Factor X is replaced with asparagine. The Factor X variant has substantially reduced affinity for Factor Va, although the catalytic impact of Factor Va binding

remains essentially intact.

IT **212622-93-0**
RL: BPR (Biological process); BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study); PROC (Process)
(amino acid sequence; factor X substitution mutant with reduced affinity for factor Va)

REFERENCE COUNT: 6 THERE ARE 6 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 12 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1998:608718 CAPLUS
DOCUMENT NUMBER: 129:213519
TITLE: Factor X deletion mutants with modified protease cleavage sites for use in hemostatics
INVENTOR(S): Himmelspach, Michele; Pfleiderer, Michael; Falkner, Falko-gunter; Eibl, Johann; Dorner, Friedrich; Schlokat, Uwe
PATENT ASSIGNEE(S): Immuno A.-G., Austria
SOURCE: PCT Int. Appl., 82 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: German
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9838318	A1	19980903	WO 1998-AT46	19980227
W: AU, BR, CA, CZ, HU, IL, JP, MX, NO, PL, RU, SI, SK, US				
RW: AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE				
AT 9700336	A	19990115	AT 1997-336	19970227
AT 405517	B	19990927		
AU 9860808	A1	19980918	AU 1998-60808	19980227
AU 732953	B2	20010503		
BR 9807618	A	20000215	BR 1998-7618	19980227
EP 1012303	A1	20000628	EP 1998-905134	19980227
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, FI				
JP 2001513632	T2	20010904	JP 1998-537063	19980227
NO 9904136	A	19991027	NO 1999-4136	19990826
US 6562598	B1	20030513	US 1999-367777	19991118
US 2003138914	A1	20030724	US 2003-348504	20030121

PRIORITY APPLN. INFO.:

AT 1997-336	A	19970227
WO 1998-AT46	W	19980227
US 1999-367777	A3	19991118

ED Entered STN: 25 Sep 1998
AB The invention relates to factor X.DELTA. analogs, comprising a deletion of the amino acids Arg180 to Arg234 and a modification in the region of the amino acid sequence between Gly173 and Arg179. These modified cleavage sites prevent proteolysis by endogenous proteases and facilitate controlled cleavage of the precursor. The invention also relates to prepns. contg. said factor X.DELTA. analogs and methods for the prodn. thereof. The prepro-factor X analogs may be used to produce high-purity factor X for use in hemostatics.

IT **103850-93-7DP**, sequence variants
RL: BPN (Biosynthetic preparation); BPR (Biological process); BSU (Biological study, unclassified); PUR (Purification or recovery); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); PROC (Process); USES (Uses)
(amino acid sequence; factor X deletion mutants with modified protease cleavage sites for use in hemostatics)

REFERENCE COUNT: 2 THERE ARE 2 CITED REFERENCES AVAILABLE FOR THIS

RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 13 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1998:608717 CAPLUS
DOCUMENT NUMBER: 129:213518
TITLE: Factor X analogs with a modified protease cleavage site
INVENTOR(S): Himmelspach, Michele; Schlokat, Uwe; Dorner, Friedrich; Fisch, Andreas; Eibl, Johann
PATENT ASSIGNEE(S): Immuno A.-G., Austria
SOURCE: PCT Int. Appl., 86 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: German
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9838317	A1	19980903	WO 1998-AT45	19980227
W: AU, BR, CA, CZ, HU, IL, JP, MX, NO, PL, RU, SI, SK, US				
RW: AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE				
AT 9700335	A	19990115	AT 1997-335	19970227
AT 405516	B	19990927		
AU 9862002	A1	19980918	AU 1998-62002	19980227
AU 744428	B2	20020221		
EP 966536	A1	19991229	EP 1998-903943	19980227
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, FI				
BR 9807627	A	20000222	BR 1998-7627	19980227
JP 2001513631	T2	20010904	JP 1998-537062	19980227
SK 282369	B6	20020107	SK 1999-1170	19980227
MX 9907768	A	20000831	MX 1999-7768	19990823
NO 9904139	A	19991027	NO 1999-4139	19990826
US 6573071	B1	20030603	US 1999-367791	19991112
US 2003181381	A1	20030925	US 2003-407123	20030404

PRIORITY APPLN. INFO.:

AT 1997-335	A	19970227
AT 1997-336	A	19970227
WO 1998-AT45	W	19980227
US 1999-367791	A3	19991112

ED Entered STN: 25 Sep 1998
AB The invention relates to factor X analogs which have a modification in the area of the naturally occurring factor Xa activating cleavage site, said modification representing a processing site of a protease which does not naturally cleave in this area of the factor X sequence. These modified cleavage sites prevent proteolysis by endogenous proteases and facilitate controlled cleavage of the precursor. The invention also relates to prepns. contg. the innovative factor X analogs and to methods for the prodn. thereof. The prepro-factor X analogs may be used to produce high-purity factor X for use in hemostatics.
IT **103850-93-7DP**, sequence variants **212379-48-1P** **212380-18-2DP**, sequence variants **212380-22-8DP**, sequence variants
RL: BPN (Biosynthetic preparation); BPR (Biological process); BSU (Biological study, unclassified); PUR (Purification or recovery); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); PROC (Process); USES (Uses)
(amino acid sequence; factor X analogs with modified protease cleavage site)
REFERENCE COUNT: 2 THERE ARE 2 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 14 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1996:161257 CAPLUS
DOCUMENT NUMBER: 124:197097
TITLE: Analogs of blood-coagulation factors for regulation of thrombus formation
INVENTOR(S): Wolf, David L.; Sinha, Uma
PATENT ASSIGNEE(S): Cor Therapeutics, Inc., USA
SOURCE: PCT Int. Appl., 69 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 3
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9600577	A1	19960111	WO 1995-US8368	19950628
W: AU, CA, JP, MX				
RW: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE				
US 5583107	A	19961210	US 1994-268003	19940629
AU 9529585	A1	19960125	AU 1995-29585	19950628
AU 712271	B2	19991104		
EP 766563	A1	19970409	EP 1995-925461	19950628
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE				
JP 10502351	T2	19980303	JP 1996-503483	19950628

PRIORITY APPLN. INFO.:

US 1994-268003	A	19940629
US 1990-578646	A2	19900904
US 1991-808329	B1	19911216
US 1994-249777	A2	19940526
WO 1995-US8368	W	19950628

ED Entered STN: 20 Mar 1996
AB Analogs of blood coagulation factors that can be used in the treatment or prevention of thrombosis or hemophilia are described. Inactive analogs of clotting factors that can prevent the formation of an active prothrombinase complex and thrombus formation are described for prevention of thrombosis. These inactive analogs may be either transiently or permanently inactive. Modified forms of activated coagulation factors that have extended half-lives are useful in treating hemophilic conditions. In particular, analogs of factor Xa are described. One form is a catalytically inactive two-chain analog and a second is a truncated analog that can be proteolytically processed but that is inactive after processing. Transient inactivation can be brought about by acylation of the factor with reactivation occurring after slow in vivo deacylation.
IT **174294-61-2P 174294-62-3P 174294-63-4P**
RL: BPN (Biosynthetic preparation); PRP (Properties); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES (Uses)
(amino acid sequence; analogs of blood-coagulation factors for regulation of thrombus formation)
IT **103850-95-9DP**, analogs
RL: BPN (Biosynthetic preparation); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES (Uses)
(amino acid sequence; analogs of blood-coagulation factors for regulation of thrombus formation)

L7 ANSWER 15 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1992:81978 CAPLUS
DOCUMENT NUMBER: 116:81978
TITLE: Structurally homologous ligand binding of integrin Mac-1 and viral glycoprotein C receptors
AUTHOR(S): Altieri, Dario C.; Etingin, Orli R.; Fair, Daryl S.; Brunck, Terence K.; Geltosky, John E.; Hajjar, David P.; Edgington, Thomas S.

CORPORATE SOURCE: Dep. Immunol., Scripps Res. Inst., La Jolla, CA, 92037, USA
SOURCE: Science (Washington, DC, United States) (1991), 254(5035), 1200-2
CODEN: SCIEAS; ISSN: 0036-8075
DOCUMENT TYPE: Journal
LANGUAGE: English
ED Entered STN: 06 Mar 1992
AB Three spatially distant surface loops were found to mediate the interaction of the coagulation protein factor X with the leukocyte integrin Mac-1. This interacting region, which by computational modeling defines a 3-dimensional macromotif in the catalytic domain, was also recognized by glycoprotein C (gC), a factor X receptor expressed on herpes simplex virus (HSV)-infected endothelial cells. Peptidyl mimicry of each loop inhibited factor X-binding to Mac-1 and gC, blocked monocyte generation of thrombin, and prevented monocyte adhesion to HSV-infected endothelium. These data link the ligand recognition of Mac-1 to established mechanisms of receptor-mediated vascular injury.
IT **138824-08-5**
RL: BIOL (Biological study)
(of blood coagulation factor X, Mac-1 and viral glycoprotein C binding sites in relation to)

L7 ANSWER 16 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1986:528609 CAPLUS
DOCUMENT NUMBER: 105:128609
TITLE: Gene for human factor X: a blood coagulation factor whose gene organization is essentially identical with that of factor IX and protein C
AUTHOR(S): Leytus, Steven P.; Foster, Donald C.; Kurachi, Kotoku; Davie, Earl W.
CORPORATE SOURCE: Dep. Biochem., Univ. Washington, Seattle, WA, 98195, USA
SOURCE: Biochemistry (1986), 25(18), 5098-102
CODEN: BICHAW; ISSN: 0006-2960
DOCUMENT TYPE: Journal
LANGUAGE: English
ED Entered STN: 18 Oct 1986
AB Factor X [9001-29-0] is 1 of 6 vitamin K-dependent proteins known to be involved in blood coagulation, the others being factor VII, factor IX [9001-28-9], prothrombin, protein S, and protein C. Recombinant phages contg. overlapping DNA inserts coding for the gene for human factor X were isolated and characterized. These DNA inserts code for almost the entire gene for factor X, extending from the prepro leader peptide through the 3'-noncoding region of the transcription product. The organization of the gene for factor X was established by DNA sequencing to identify the location of the introns and exons in the gene. Seven introns and 8 exons were identified and their intron/exon boundaries established. The 7 introns interrupt the coding sequence at essentially identical locations in the amino acid sequence, as do the introns in the genes for human factor IX and protein C. The introns in the gene for factor X divide the coding sequence into discrete exons that code for potential structural and functional domains of the protein. Apparently, the vitamin K-dependent proteins present in plasma have evolved from a single, common gene, and this ancestral gene arose through a process that involved the assembly of small protein coding units of DNA into a single gene.
IT **103850-91-5 103850-92-6 103850-93-7**
103850-94-8 103850-95-9
RL: PRP (Properties)
(amino acid sequence of)

L7 ANSWER 17 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN

ACCESSION NUMBER: 1986:492292 CAPLUS
DOCUMENT NUMBER: 105:92292
TITLE: Isolation and characterization of human blood-coagulation factor X cDNA
AUTHOR(S): Kaul, Rajinder K.; Hildebrand, Beth; Roberts, Savithri; Jagadeeswaran, Pudur
CORPORATE SOURCE: Coll. Med., Univ. Illinois, Chicago, IL, 60612, USA
SOURCE: Gene (1986), 41(2-3), 311-14
CODEN: GENED6; ISSN: 0378-1119
DOCUMENT TYPE: Journal
LANGUAGE: English
ED Entered STN: 19 Sep 1986
AB Synthetic oligodeoxynucleotides were used as probes, to isolate factor X [9001-29-0] cDNA from a human liver cDNA library. The 1430-base-pair cDNA which spans the coding region of the mature factor X and contains the polyadenylation signal and poly(A) tail was sequenced. The amino acid (aa) sequence is in agreement with the published aa sequence. The nucleotide (nt) sequence of cDNA confirmed that factor X is synthesized and secreted as a single-chain precursor, and then converted into dimeric form by proteolytic cleavage of an internal tripeptide. From the nt sequence, it was also predicted that like other secretory proteins, human factor X is synthesized with a leader sequence (prepro-protein). The 5'-coding region of factor X cDNA is 60 and 40% homologous to the corresponding regions of factor IX [9001-28-9] and prothrombin [9001-26-7] genes, resp. This supports the hypothesis of gene evolution by gene duplication followed by divergence.
IT **103813-13-4**
RL: PRP (Properties)
(amino acid sequence of)

L7 ANSWER 18 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1985:482593 CAPLUS
DOCUMENT NUMBER: 103:82593
TITLE: Characterization of an almost full-length cDNA coding for human blood coagulation factor X
AUTHOR(S): Fung, Marion R.; Hay, Colin W.; MacGillivray, Ross T. A.
CORPORATE SOURCE: Dep. Biochem., Univ. Br. Columbia, Vancouver, BC, V6T 1W5, Can.
SOURCE: Proceedings of the National Academy of Sciences of the United States of America (1985), 82(11), 3591-5
CODEN: PNASA6; ISSN: 0027-8424
DOCUMENT TYPE: Journal
LANGUAGE: English
ED Entered STN: 22 Sep 1985
AB A human liver cDNA library was screened by colony hybridization with a bovine factor X cDNA probe. Three of the pos. plasmids contained overlapping DNA that coded for most of human factor X [9001-29-0] mRNA. DNA sequence anal. of these 3 clones allowed the prediction of the complete amino acid sequence of plasma factor X. From these studies, it was possible to predict that human factor X is synthesized as a single polypeptide chain precursor in which the light and heavy chains of plasma factor X are linked by the tripeptide Arg-Lys-Arg. The cDNA sequence also predicts that human factor X is synthesized as a preproprotein having an N-terminal leader peptide of .gtoreq.28 amino acid residues. A comparison of the amino acid sequences of human and bovine factor X shows high sequence identity around the Ca-binding regions and catalytic regions but low sequence identity around the nonfunctional regions.
IT **91930-64-2 91930-66-4**
RL: PRP (Properties)
(amino acid sequence of)

L7 ANSWER 19 OF 26 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1984:524051 CAPLUS
DOCUMENT NUMBER: 101:124051
TITLE: Characterization of a cDNA coding for human factor X
AUTHOR(S): Leytus, Steven P.; Chung, Dominic W.; Kisiel, Walter; Kurachi, Kotoku; Davie, Earl W.
CORPORATE SOURCE: Dep. Biochem., Univ. Washington, Seattle, WA, 98195, USA
SOURCE: Proceedings of the National Academy of Sciences of the United States of America (1984), 81(12), 3699-702
CODEN: PNASA6; ISSN: 0027-8424
DOCUMENT TYPE: Journal
LANGUAGE: English
ED Entered STN: 13 Oct 1984
AB A phage .lambda.gt11 cDNA library contg. DNA inserts prepd. from human liver mRNA was screened with an antibody to human blood coagulation factor X [**91930-66-4**], a plasma protein participating in the middle phase of the blood coagulation cascade. Ten pos. clones were isolated from 2 .times. 106 phages and plaque purified. The cDNA in the phage contg. the largest insert was sequenced and shown to code for human factor X. This cDNA insert contained 1137 base pairs (bp) coding for a portion of the light chain of the mol., a connecting region, the heavy chain, a stop codon, a short 3' noncoding region, the heavy chain, a stop codon, a short 3' noncoding region, and a poly(A) tail. The sequence of A-T-T-A-A-A, which functions as a potential recognition site for polyadenylylation or processing, was present in the 3' end of the coding sequence and preceded the stop codon of TGA by 1 bp and the poly(A) tail by 14 bp. The amino acid sequence deduced from the cDNA indicated that factor X is synthesized as a single-chain polypeptide contg. the light and heavy chains connected by an Arg-Lys-Arg tripeptide. The single-chain mol. is then converted to the light and heavy chains by cleavage of .gtoreq.2 internal peptide bonds. In plasma, these 2 chains are linked together by an SS bond. The DNA sequence coding for the active site of human factor X showed a high degree of identify with prothrombin and factor IX, 2 other vitamin K-dependent serine proteases that participate in blood coagulation. These data, along with the protein sequence data previously published for the light chain of human factor X, establish the complete amino acid sequence for the mature protein present in plasma.
IT **91930-64-2 91930-66-4**
RL: PRP (Properties)
(amino acid sequence of)

L7 ANSWER 20 OF 26 USPATFULL on STN
ACCESSION NUMBER: 2004:57015 USPATFULL
TITLE: Prothrombin activating protein
INVENTOR(S): Masci, Paul Pantaleone, Brisbane, AUSTRALIA
Jersey, John De, Brisbane, AUSTRALIA
Lavin, Martin, Brisbane, AUSTRALIA

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 2004043017	A1	20040304	
APPLICATION INFO.:	US 2003-406031	A1	20030402	(10)

	NUMBER	DATE
PRIORITY INFORMATION:	AU 2002-1483	20020403
	AU 2003-2003901033	20030307

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: FISH & RICHARDSON PC, 225 FRANKLIN ST, BOSTON, MA, 02110

NUMBER OF CLAIMS: 56
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 47 Drawing Page(s)
LINE COUNT: 4925
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
AB The invention relates to snake venom protease polypeptides and nucleic acid sequences encoding same. This invention also relates to methods of making and using the snake venom proteases, e.g., to promote haemostasis and prevent blood loss such as during surgery or for treatment of wounds resulting from accidents and other types of injury or trauma.
IT **612559-84-9**
(unclaimed sequence; prothrombin activating proteins SVPs (snake venom proteases) from Australian snakes, nucleic acid sequences encoding same, and methods to promote hemostasis and prevent blood loss)

L7 ANSWER 21 OF 26 USPATFULL on STN
ACCESSION NUMBER: 2004:13030 USPATFULL
TITLE: Export and modification of (poly)peptides in the lantibiotic way
INVENTOR(S): Moll, Gert Nikolaas, Groningen, NETHERLANDS
Leenhouts, Cornelis Johannes, Haren, NETHERLANDS

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 2004009550	A1	20040115	
APPLICATION INFO.:	US 2003-360101	A1	20030207	(10)

	NUMBER	DATE
PRIORITY INFORMATION:	EP 2002-77060	20020524

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: TRASK BRITT, P.O. BOX 2550, SALT LAKE CITY, UT, 84110
NUMBER OF CLAIMS: 13
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 9 Drawing Page(s)
LINE COUNT: 3337
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
AB The invention includes a method for harvesting a polypeptide produced by a host cell, wherein the polypeptide has not undergone intra-cellular post-translational modification, such as dehydration of a serine or a threonine, and/or thioether bridge formation. The invention also includes a method for producing thioether containing peptides and dehydroalanine/dehydrobutyrine-containing peptides, wherein extracellularly thioether rings may be formed.
IT **631960-24-2**
(unclaimed sequence; use of the lantibiotic transport system to secrete foreign proteins into culture medium for purifn.)

L7 ANSWER 22 OF 26 USPATFULL on STN
ACCESSION NUMBER: 2003:318639 USPATFULL
TITLE: Atherosclerotic phenotype determinative genes and methods for using the same
INVENTOR(S): West, Mike, Durham, NC, UNITED STATES
Nevins, Joseph R., Chapel Hill, NC, UNITED STATES
Goldschmidt, Pascal, Chapel Hill, NC, UNITED STATES

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 2003224383	A1	20031204	
APPLICATION INFO.:	US 2002-291885	A1	20021112	(10)

	NUMBER	DATE
PRIORITY INFORMATION:	US 2002-374547P	20020423 (60)
	US 2002-420784P	20021024 (60)
	US 2002-421043P	20021025 (60)
	US 2002-424680P	20021108 (60)

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: Gregory J. Glover, Ropes & Gray, Suite 800 East, 1301 K Street, NW, Washington, DC, 20005
NUMBER OF CLAIMS: 29
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 235 Drawing Page(s)
LINE COUNT: 2165
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB Genes whose expression is correlated with and determinant of an atherosclerotic phenotype are provided. Also provided are methods of using the subject atherosclerotic determinant genes in diagnosis and treatment methods, as well as drug screening methods. In addition, reagents and kits thereof that find use in practicing the subject methods are provided. Also provided are methods of determining whether a gene is correlated with a disease phenotype, where correlation is determined using at least one parameter that is not expression level and is preferably determined using a Bayesian analysis.

IT **481289-90-1**, GenBank AAA52764
(amino acid sequence; genes expressed in atherosclerotic tissue and their use in diagnosis and pharmacogenetics)

L7 ANSWER 23 OF 26 USPATFULL on STN
ACCESSION NUMBER: 2003:258328 USPATFULL
TITLE: Factor X analogues having a modified protease cleavage site
INVENTOR(S): Himmelspach, Michele, Leopoldsdorf, AUSTRIA
Schlokat, Uwe, Orth/Donau, AUSTRIA
Dorner, Friedrich, Vienna, AUSTRIA
Fisch, Andreas, St. Gallen, SWITZERLAND
Eibl, Johann, Vienna, AUSTRIA

	NUMBER	KIND	DATE
PATENT INFORMATION:	US 2003181381	A1	20030925
APPLICATION INFO.:	US 2003-407123	A1	20030404 (10)

RELATED APPLN. INFO.: Division of Ser. No. US 1999-367791, filed on 12 Nov 1999, GRANTED, Pat. No. US 6573071 A 371 of International Ser. No. WO 1998-AT45, filed on 27 Feb 1998, UNKNOWN

	NUMBER	DATE
PRIORITY INFORMATION:	AT 1997-335	19970227

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: TOWNSEND AND TOWNSEND AND CREW, LLP, TWO EMBARCADERO CENTER, EIGHTH FLOOR, SAN FRANCISCO, CA, 94111-3834
NUMBER OF CLAIMS: 51
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 13 Drawing Page(s)
LINE COUNT: 2349
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB Factor X analogues having a modification in the region of the natural Factor Xa activation cleavage site, said modification representing a processing site of a protease not naturally cleaving in this region of

the Factor X sequence, preparations containing the Factor X analogues according to the invention, and processes for the preparation thereof are described.

IT **103850-93-7DP**, sequence variants **212379-48-1P** **212380-18-2DP**, sequence variants **212380-22-8DP**, sequence variants
(amino acid sequence; factor X analogs with modified protease cleavage site)

L7 ANSWER 24 OF 26 USPATFULL on STN

ACCESSION NUMBER:	2003:200929 USPATFULL
TITLE:	Factor X deletion mutants and analogues thereof
INVENTOR(S):	Himmelspach, Michele, Leopoldsdorf, AUSTRIA Pfleiderer, Michael, Darmstadt, GERMANY, FEDERAL REPUBLIC OF Falkner, Falko-Guenter, Orth/Donau, AUSTRIA Eibl, Johann, Vienna, AUSTRIA Dorner, Friedrich, Vienna, AUSTRIA Schlokat, Uwe, Orth/Donau, AUSTRIA

	NUMBER	KIND	DATE
PATENT INFORMATION:	US 2003138914	A1	20030724
APPLICATION INFO.:	US 2003-348504	A1	20030121 (10)

RELATED APPLN. INFO.: Division of Ser. No. US 1999-367777, filed on 18 Nov 1999, GRANTED, Pat. No. US 6562598 A 371 of International Ser. No. WO 1998-AT46, filed on 27 Feb 1998, UNKNOWN

	NUMBER	DATE
PRIORITY INFORMATION:	AT 1997-336	19970227

DOCUMENT TYPE:	Utility
FILE SEGMENT:	APPLICATION
LEGAL REPRESENTATIVE:	TOWNSEND AND TOWNSEND AND CREW, LLP, TWO EMBARCADERO CENTER, EIGHTH FLOOR, SAN FRANCISCO, CA, 94111-3834
NUMBER OF CLAIMS:	45
EXEMPLARY CLAIM:	1
NUMBER OF DRAWINGS:	12 Drawing Page(s)
LINE COUNT:	2232

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB Factor X.DELTA. analogues having a deletion of amino acids Arg180 to Arg234 and a modification in the region of the amino acid sequence between Gly173 and Arg179, preparations containing these factor X.DELTA. analogues, and processes for the preparation thereof are described.

IT **103850-93-7DP**, sequence variants **212379-48-1P** **212380-18-2DP**, sequence variants **212380-22-8DP**, sequence variants
(amino acid sequence; factor X analogs with modified protease cleavage site)

L7 ANSWER 25 OF 26 USPATFULL on STN

ACCESSION NUMBER:	2003:148878 USPATFULL
TITLE:	Factor X analogues with a modified protease cleavage site
INVENTOR(S):	Himmelspach, Michele, Leopoldsdorf, AUSTRIA Schlokat, Uwe, Orth/Donau, AUSTRIA Dorner, Friedrich, Vienna, AUSTRIA Fisch, Andreas, St. Gallen, SWITZERLAND Eibl, Johann, Vienna, AUSTRIA
PATENT ASSIGNEE(S):	Baxter Aktiengesellschaft, Vienna, AUSTRIA (non-U.S. corporation)

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 6573071	B1	20030603	
	WO 9838317		19980903	
APPLICATION INFO.:	US 1999-367791		19991112	(9)
	WO 1998-AT45		19980227	

	NUMBER	DATE
PRIORITY INFORMATION:	AT 1997-335	19970227

DOCUMENT TYPE: Utility
FILE SEGMENT: GRANTED
PRIMARY EXAMINER: Low, Christopher S. F.
ASSISTANT EXAMINER: Schnizer, Holly
LEGAL REPRESENTATIVE: Townsend and Townsend and Crew, L.L.P.
NUMBER OF CLAIMS: 64
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 13 Drawing Figure(s); 13 Drawing Page(s)
LINE COUNT: 2472
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB Factor X analogues having a modification in the region of the natural Factor Xa activation cleavage site, said modification representing a processing site of a protease not naturally cleaving in this region of the Factor X sequence, preparations containing the Factor X analogues according to the invention, and processes for the preparation thereof are described.

IT **103850-93-7DP**, sequence variants **212379-48-1P**
212380-18-2DP, sequence variants **212380-22-8DP**, sequence variants
(amino acid sequence; factor X analogs with modified protease cleavage site)

L7 ANSWER 26 OF 26 USPATFULL on STN
ACCESSION NUMBER: 2003:129813 USPATFULL
TITLE: Factor X deletion mutants and analogues thereof
INVENTOR(S): Himmelspach, Michele, Leopoldsdorf, AUSTRIA
Pfleiderer, Michael, Darmstadt, GERMANY, FEDERAL REPUBLIC OF
Falkner, Falko-Guenter, Orth/Donau, AUSTRIA
Eibl, Johann, Vienna, AUSTRIA
Dorner, Friedrich, Vienna, AUSTRIA
Schlokat, Uwe, Orth/Donau, AUSTRIA
PATENT ASSIGNEE(S): Baxter Aktiengesellschaft, Vienna, AUSTRIA (non-U.S. corporation)

	NUMBER	KIND	DATE	
PATENT INFORMATION:	US 6562598	B1	20030513	
	WO 9838318		19980903	
APPLICATION INFO.:	US 1999-367777		19991118	(9)
	WO 1998-AT46		19980227	

	NUMBER	DATE
PRIORITY INFORMATION:	AU 1997-336	19970227

DOCUMENT TYPE: Utility
FILE SEGMENT: GRANTED
PRIMARY EXAMINER: Carlson, Karen Cochrane
ASSISTANT EXAMINER: Snedden, Sheridan
LEGAL REPRESENTATIVE: Townsend and Townsend and Crew LLP
NUMBER OF CLAIMS: 56

EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 15 Drawing Figure(s); 12 Drawing Page(s)
LINE COUNT: 2334
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB Factor X.DELTA. analogues are provided, as well as pharmaceutical preparations containing such analogues and methods of preparing such analogues. The factor X.DELTA. analogues have a deletion of the amino acids Arg180 to Arg234 and a modification in the region of the amino acid sequence between Gly173 and Arg179 of the factor X amino acid sequence. Such analogues can include a processing site not normally present in factor X, thus allowing for selective conversion of the analogue to an active form. The analogues and preparations have utility in the treatment of a number of blood coagulation disorders.

IT **103850-93-7DP**, sequence variants **212379-48-1P**
212380-18-2DP, sequence variants **212380-22-8DP**,
sequence variants
(amino acid sequence; factor X analogs with modified protease cleavage site)

FILE 'HOME' ENTERED AT 14:07:37 ON 05 OCT 2004